译者介绍

庄兴无 福建师范大学教授, 1962 年福建师范学院数学专业本科毕业. 1979—1981 年赴罗马尼亚布加勒斯特大学数学系做访问学者. 1994 年赴美国密苏里大学数学系做高级访问学者. 从事概率论与数理统计研究, 在国内外杂志发表学术论文三十多篇. 曾主持两项国家自然科学基金项目和多项福建省科技研究项目, 获福建省科技进步二等奖、福建省高校科技成果三等奖、福建省教学成果二等奖.

陈宗洵 福建师范大学教授, 福建省统计局特约研究员. 1963 年福建师范学院数学专业本科毕业. 从事应用概率统计研究, 发表学术论文约 20 篇, 涉及多学科、多领域. 获福建省科技进步二等奖 (排名第二)、三等奖 (子课题负责人).

陈庆华 福建师范大学副教授, 1983 年福建师范大学数学专业本科毕业, 1986 年福建师范大学概率论与数理统计专业硕士研究毕业, 留校任教至今, 硕士生导师. 2006 年上海大学运筹学与控制论专业博士研究生毕业, 获理学博士学位. 从事应用概率统计研究, 研究方向为随机图与复杂网络、可靠性数学, 已在《PHYSICAL REVIEWE》等国内外学术刊物上发表论文数篇. 目前, 主持一项福建省自然科学基金项目和一项福建省科技计划项目, 参加一项国家自然科学基金项目.

推 荐 序

人民邮电出版社图灵公司数学和统计方向的策划编辑, 曾就引进国外优秀的随机过程方向教科书征询过我的意见. 后来他们决定翻译出版 S. Karlin 和 H. M. Taylor 的*A First Course in Stochastic Processes*, 并邀请我担任此书的译者. 我告诉他们, 早在 20 世纪 80 年代, 福建师范大学数学系庄兴无教授等就已经翻译过这本教材的绝大部分. 然后, 通过我的博士生 (福建师范大学数学系陈庆华博士) 促成了此事. 在本书中译本出版前夕, 他们三位译者要我代写一篇序言来介绍这本教科书的特色.

S. Karlin 是国际上著名的应用概率专家, 斯坦福大学数学系荣休教授. 生灭过程中计算平稳分布的 Karlin-McGregor 定理是众所周知的. 他的博士生 M. F. Neuts 教授作为*Stochastic Models* 杂志的奠基人, 通过引入 PH 分布创立了矩阵解析方法, 在拟生灭过程方面发扬光大了他的工作. 此外, Karlin 教授在生物信息学方面也有许多重要的建树. 我和 Neuts 教授有 20 多年的交往, 在我们 2006 年出版的《生物信息学 —— 智能化算法及其应用》一书中就引用了 Karlin 教授的一项研究成果.

我认为 Karlin 和 Taylor 著的《随机过程初级教程》是应用随机过程方面值得引进和借鉴的一本优秀入门教材, 它的最大特点是"理论与应用紧密地结合在一起, 使两者相得益彰", 这也是作者写作本书的宗旨.

Karlin 和 Taylor 著的随机过程分为两卷: 初级教程和高级教程. 现在翻译出版的是初级教程, 包括了马尔可夫链、更新过程、鞅过程、布朗运动、分支过程和平稳过程等, 涵盖了应用随机过程中最重要、最常用的过程类型. 离散时间马尔可夫链和连续时间马尔可夫链作为应用随机过程的基础内容, 作者用了三章篇幅进行论述. 因为"鞅的概念和方法对于分析各类随机过程泛函是极为重要的工具", 作者在第 2 版特意增加了这一章. 又如更新现象在排队论、可靠性理论和库存论中非常普遍, 而布朗运动和分支过程在物理学、生物学、金融学和社会科学中有着广泛的应用, 所以都各有一章讨论. 此外, 有些随机过程教材很少涉及平稳过程, 但本书专门开辟了一章介绍平稳过程, 使得内容更为完整.

由于随机过程应用的广泛性, 采用较多的实际例子来说明概念和结论是十分流行的写作方法. 但是严谨的理论论述和众多的实例解释交织在一起, 如果处理不好就会显得杂乱无章, 反而让读者一头雾水. 然而, 本书在这个关键问题上平衡得恰到好处, 能将理论与应用有机地结合在一起, 读起来特别流畅. 这样, 使学习纯数学的学生知道随机过程的多种应用, 增加学习的兴趣; 同时, 也使从事应用数学的学生认识到基础理论在实际工作中的重要性.

本书另外一个特色是每章都附有大量的习题. 习题分为初等的和较难的两部

分, 这就使得读者可根据自己的水平有所选择. 做习题不仅可以加深对所学内容的理解, 而且还能够通过较难的习题扩充相关理论. 要掌握随机过程的理论, 并准确地应用于解决实际问题, 做习题是个不能忽视的手段.

总之, 这是一本在内容的取舍和论述的方式上都精心设计的随机过程优秀入门教材, 它将让读者迅速掌握随机过程的精髓.

史定华

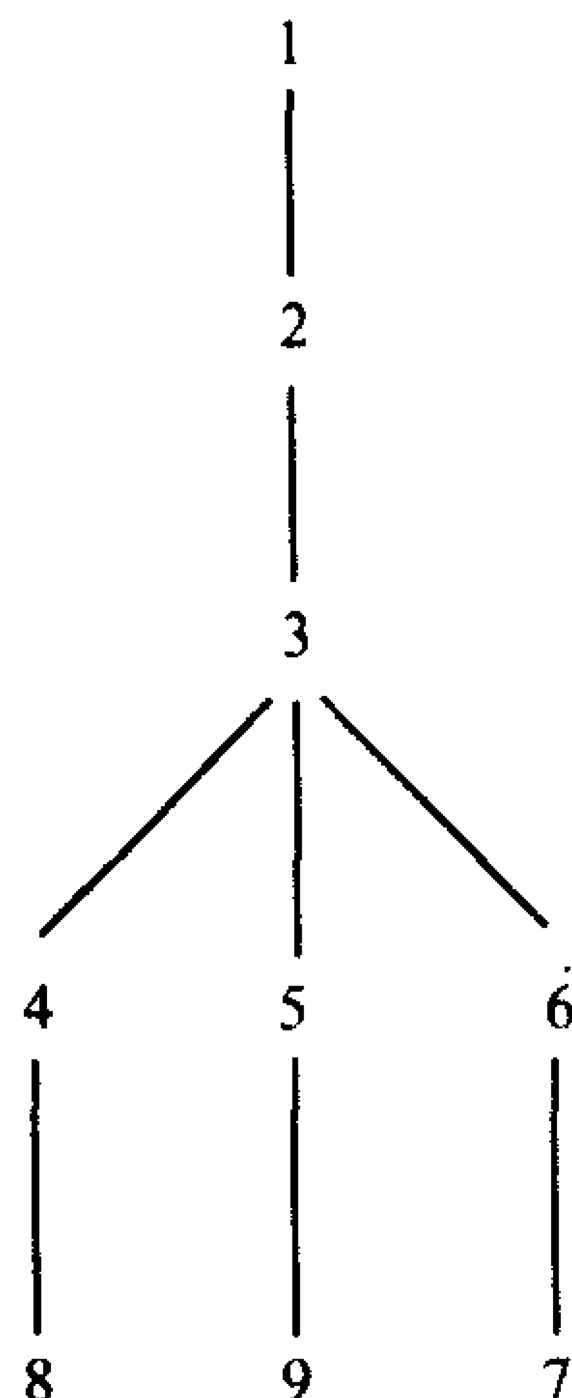
2007 年 2 月 18 日于上海

序 言

新版的目的、层次和风格与原版所提出的宗旨是一致的. 我们仍尽力把理论和应用紧密结合在一起, 使两者相得益彰.

新版主要做了三个方面的变动. 首先, 我们充实了第 1 版中的若干专题. 其次, 在每一章末增加了很多练习和问题. 第三方面, 也是最主要的变动, 我们增加了新的章节, 扼要讨论了第 1 版未涉及的几种随机过程, 如: 鞅、与随机和相联系的更新和波动现象、平稳随机过程以及扩散理论.

鞅的概念和方法对于分析多种类型随机过程泛函是极为重要的工具. 特别地, 鞅模型已有效地用于扩散型随机模型的研究. 更新现象在工程学和管理科学中, 尤其在可靠性、排队和存储系统中几乎是同样重要的. 我们将增加一章来系统地讨论更新理论, 新增的另一章将探讨平稳过程理论及其在某些工程和经济问题中的应用, 其他的新章节将研究可以用来描述某些生物和物理系统的扩散过程理论, 以及在分析排队系统和运筹学某些方面颇为有用的独立随机变量和的波动性质.



各章之间逻辑上的联系由右图说明. 1.1 节可浏览而过, 不必拘泥于细节. 第 7 章中只有 7.5 节和 7.6 节与第 6 章有关. 第 9 章中只有 9.9 节与第 5 章有关.

第 1 章的 1.2 节和 1.3 节 (1.1 节的内容只须粗略地复习)、第 2 章、第 3 章 (除了 3.5 节和 3.6 节)、第 4 章 (除了 4.3 节、4.7 节和 4.8 节) 等部分可以组成适合于大学三、四年级一个学期课程的教材. 至于后面内容的选取可由讲授人斟酌确定, 从第 5 章至第 9 章的前面几节选取一些材料比较适宜.

章末的问题分成两组, 第一组比较初等, 第二组比较困难和精细.

本书涉及范围十分广泛, 因此分为 "初级教程" 和 "高级教程" 两卷. 初级教程包括作为理论基础并与应用关系最密切的几种主要随机过程. 在高级教程中我们将介绍另外一些论题和应用, 并更深入研究初级教程中提出的一些问题. 我们选材考虑到要适应更广泛的读者, 包括从事有关数学的、工程的、物理学的、生物学的、社会科学以及管理科学中随机分析的理论研究者和实际工作者.

作为本书下卷的《随机过程高级教程》将包括如下几章: 第 10 章, 马尔可夫链的代数方法; 第 11 章, 转移概率的比定理及应用; 第 12 章, 作为马尔可夫链的独立随机变量和; 第 13 章, 顺序统计量、泊松过程和应用; 第 14 章, 连续时间马尔可夫链; 第 15 章, 扩散过程; 第 16 章, 复合随机过程; 第 17 章, 独立同分布随机变量部

分和的波动理论; 第 18 章, 排队过程.

如第 1 版序言所说, 我们引用了随机过程理论和应用方面的大量文献. 每章的末尾附有某些有代表性的参考书目, 这对进一步查阅资料是有益的.

我们感谢魏兹曼科学所、斯坦福大学和康奈尔大学提供了一个资料丰富和工作方便的环境. 第一作者感谢海军研究所持续给予的大力支持, 使得我们能集中精力考虑本书的若干思想和写作计划. 我们衷心感谢同事们提出许多建设性的意见, 尤其是 La Trobe 大学的 P.Brockwell 教授、牛津大学的 J. Kingman 教授、斯坦福大学的 D. Iglehart 和 S.Ghurye 教授以及康奈尔大学的伊藤清教授和 S. Stidham, Jr. 教授. 我们也感谢我们的学生 M. Nedzela 和 C. Macken 协助检查习题并校对文字.

Samuel Karlin
Howard M. Taylor

第 1 版序言

随机过程主要考虑服从概率规律的事件序列. 它广泛应用于物理学、工程学、生物学、医学、心理学和其他学科, 以及数学的其他分支. 本书的目的是为随机过程的许多专题提供一个导引. 特别地, 我力图做到以下三点: (1) 提出随机过程几个主要领域的系统性导引; (2) 使纯数学专业的学生对随机过程的多种应用产生兴趣; (3) 使偏重于应用的学生认识到随机过程的基础数学理论的重要性.

本书的例子主要来源于生物学和工程学, 但着重于具有数学趣味的或者在若干学科中都有重要意义的随机模型. 我们要讨论和阐述概率论中的许多著名的概念和问题.

由于在一本初级教程中不可能讨论这个领域的所有方面, 故我们省略了一些重要的专题, 例如, 平稳随机过程和鞅. 本书在任何意义下都不希望成为一本权威性著作, 相反, 本书的目的只是希望在初等概率论和许多优秀的高级随机过程论著之间搭建一座桥梁.

我们假定读者熟悉初等概率论, 例如 Feller 的《概率论及其应用》前半部分[1]. 1.1 节介绍了必要的背景材料以及本书的术语和记号. 部分内容初读时可略过. 每章末附有练习以帮助解释和扩充理论部分.

根据实际需要本书可作为一个学期或两个学期的课程来安排.

在写这本书时, 我们参考了大量随机过程的文献. 每章末都附有相关的参考书目, 以便于进一步查阅材料和文献.

我感谢斯坦福大学和美国海军科研局为写此书提供的各种方便和经费支持. 感谢我的同事, 如斯坦福大学的钟开莱教授和 J. McGregor 教授, 他们经常给予我鼓励和帮助, 感谢达特茅斯学院的 J.Lamperti 教授、康奈尔大学的 J.Kiefer 教授和威斯康星大学的 P.Ney 教授提出许多建设性意见, 感谢 A.Feinstein 博士详细阅读了手稿的主要部分, 感谢我的学生 P.Milch、B.Singer、M.Feldman 和 B.Krishnamoorthi 提出有益的意见并协助组织习题. 最后, 我要感谢 Gail Lemmond 和 Rosemarie Stampfel.

Samuel Karlin

1. 该书第 1 卷中文版已由人民邮电出版社出版, 第 2 卷中文版也将于 2007 年由人民邮电出版社出版.

目　录

第 1 章　随机过程初步

1.1 节概述了必要的背景材料并介绍了本书使用的术语和记号. 建议读者先浏览而过, 不必深究, 待以后需要时再回头复习.

1.2 节介绍了著名的布朗运动和泊松过程, 1.3 节概括介绍了本书后面主要涉及的各类随机过程.

为了完整起见, 最后一节讨论了一般理论中的一些技术问题. 这一节在初读时可以略过.

1.1　基本术语、随机变量和分布函数性质的复习

本节简要复习概率论中基本概念和术语. 本节内容在以后各章中将时常使用而不再详细说明. 我们建议读者考虑本章末的问题, 这些问题不仅可供练习而且有助于阐明概念. 若要详细讨论这些问题, 读者可参考任何一本标准的概率论教程. (见本章末参考书目.)

我们假定读者熟悉以下概念.

(1) 实随机变量 X.

(2) X 的分布函数 F(定义为 $F(\lambda) = \Pr\{X \leqslant \lambda\}$) 和它的初等性质.

(3) 有关随机变量 X 的事件及其概率.

(4) X 的期望 $E\{X\}$ 及高阶矩 $E\{X^n\}$.

(5) 用于计算事件概率的全概率公式和贝叶斯公式.

缩写词 r.v. 表示“实随机变量”. r.v. X 称为**离散型的**, 如果存在有限或可列个不同的值 $\lambda_1, \lambda_2, \cdots$ 使得 $a_i \equiv \Pr\{X = \lambda_i\} > 0, i = 1, 2, \cdots$, 并且 $\sum_i a_i = 1$. 若对每个实数 $\lambda, \Pr\{X = \lambda\} = 0$, 则称 r.v. X 为**连续型的**. 如果存在非负函数 $p(t), -\infty < t < +\infty$, 使得 r.v. X 的分布函数由下式确定:

$$F(\lambda) = \int_{-\infty}^{\lambda} p(t)\mathrm{d}t,$$

则称 p 为 X 的概率密度. 如果 X 具有概率密度, 它必定是连续型的. 然而有例子表明, 连续型的 r.v. 不一定具有概率密度.

如果 X 是离散型 r.v., 那么它的 m **阶矩**由下式确定:

$$E[X^m]=\sum_i \lambda_i^m \Pr\{X=\lambda_i\},$$

(这里的 λ_i 如前所述) 假定此处的级数绝对收敛.

如果 X 是具有概率密度 $p(\cdot)$ 的连续型随机变量, 它的 m **阶矩**由下式确定:

$$E[X^m]=\int_{-\infty}^{\infty} x^m p(x)\mathrm{d}x,$$

假定上述积分绝对收敛.

X 的一阶矩通常称为**均值或期望**, 用 m_X 或 μ_X 表示. 如果 m_X 存在, X 的 m **阶中心矩**定义为 r.v. $X-m_X$ 的 m 阶矩. 一阶中心矩显然是零, 二阶中心矩称为 X 的**方差**, 记为 σ_X^2. 具有性质 $\Pr\{X\geqslant\nu\}\geqslant\frac{1}{2}$ 和 $\Pr\{X\leqslant\nu\}\geqslant\frac{1}{2}$ 的值 ν 称为 r.v. X 的**中位数**.

如果 X 是随机变量, g 是函数, 那么 $Y=g(X)$ 也是随机变量. 如果 X 是离散型随机变量, 具有值 $x_1,x_2,\cdots$, 那么 $g(X)$ 的期望由下式给定:

$$E[g(X)]=\sum_{i=1}^{\infty} g(x_i)\Pr\{X=x_i\}, \tag{1.1}$$

这里, 假定级数绝对收敛. 若 X 是连续型的且具有概率密度函数 p_X, 则 $g(X)$ 的期望由下式计算:

$$E[g(X)]=\int g(x)p_X(x)\mathrm{d}x. \tag{1.2}$$

包括上述离散和连续两种情形的一般公式是

2

$$E[g(X)]=\int g(x)\mathrm{d}F_X(x), \tag{1.3}$$

这里 F_X 是 r.v. X 的分布函数, 积分 (1.3) 称为 Lebesgue-Stieltjes 积分. 在这本书中我们暂不需要这种积分的知识. 但当 X 是离散型随机变量时, (1.3) 即表示 (1.1), 当 X 具有概率密度函数 p_X 时, (1.3) 即表示 (1.2).

设 $F_Y(y)=\Pr\{Y\leqslant y\}$ 表示 $Y=g(X)$ 的分布函数, 当 X 是离散型随机变量时,

$$E[Y]=\sum y_i\Pr\{Y=y_i\}=\sum g(x_i)\Pr\{X=x_i\},$$

此处 $y_i=g(x_i)$, 并且假定第二个和是绝对收敛的. 一般地有

$$E[Y]=\int y\mathrm{d}F_Y(y)=\int g(x)\mathrm{d}F_X(x). \tag{1.4}$$

如果 X 是离散型的随机变量, 则 $Y=g(X)$ 也是离散型 r.v.; 然而, 当 X 是连续型随机变量时 Y 却可能是离散型的 (读者可自己举例说明). 即使如此, 由 (1.4) 两种方式计算 $E[Y]$ 结果都是相同的.

A. 联合分布函数

给定一对 r.v.(X,Y), 它们的联合分布函数是由下式确定的二元实变量的函数

$$F(\lambda_1,\lambda_2)=F_{XY}(\lambda_1,\lambda_2)=\Pr\{X\leqslant\lambda_1,Y\leqslant\lambda_2\},$$

(倘若不会引起混淆, 通常省略下标 XY) .

函数 $F(\lambda,+\infty)\equiv\lim\limits_{\lambda_2\to\infty}F(\lambda_1,\lambda_2)$ 是一个概率分布函数, 称为 **X的边缘分布函数**. 类似地, 函数 $F(+\infty,\lambda)$ 称为 **Y的边缘分布函数**. 如果对于任意的 λ_1,λ_2 成立 $F(\lambda_1,+\infty)\cdot F(+\infty,\lambda_2)=F(\lambda_1,\lambda_2)$, 则称随机变量 X 和 Y 是**相互独立的或独立的**. 称联合分布函数 F_{XY} 具有 (联合) 概率密度, 如果存在二元实变量的非负函数 $p_{XY}(s,t)$ 使得

$$F_{XY}(\lambda_1,\lambda_2)=\int_{-\infty}^{\lambda_2}\int_{-\infty}^{\lambda_1}p_{XY}(s,t)\mathrm{d}s\mathrm{d}t$$

对任意 λ_1,λ_2 成立. 如果 X 和 Y 是独立的, 则 $p_{XY}(s,t)$ 必有形式 $p_X(s)p_Y(t)$, 其中 p_X 和 p_Y 分别是 X 和 Y 的边缘分布概率密度.

3

n 个 r.v. $X_1,X_2,\cdots,X_n$ 的联合分布函数定义为:

$$\begin{aligned}F(\lambda_1,\lambda_2,\cdots,\lambda_n)&=F_{X_1,X_2,\cdots,X_n}(\lambda_1,\lambda_2,\cdots,\lambda_n)\\&=\Pr\{X_1\leqslant\lambda_1,X_2\leqslant\lambda_2,\cdots,X_n\leqslant\lambda_n\}.\end{aligned}$$

分布函数

$$F_{X_{i_1},\cdots,X_{i_k}}(\lambda_{i_1},\cdots,\lambda_{i_k})=\lim_{\lambda_i\to\infty.i\neq i_1,\cdots,i_k}F(\lambda_1,\cdots,\lambda_n)$$

称为随机变量 $X_{i_1},X_{i_2},\cdots,X_{i_k}$ 的边缘分布.

如果对所有值 $\lambda_1,\cdots,\lambda_n,F(\lambda_1,\cdots,\lambda_n)=F_{X_1}(\lambda_1)\cdots F_{X_n}(\lambda_n)$, 则称 n 个随机变量 $X_1,X_2,\cdots,X_n$ 是 **相互独立的** 或 **独立的**.

联合分布函数 $F(\lambda_1,\lambda_2,\cdots,\lambda_n)$ 称为具有概率密度, 如果存在 n 个实变量的非负函数 $p(t_1,\cdots,t_n)$ 使得

$$F(\lambda_1,\cdots,\lambda_n)=\int_{-\infty}^{\lambda_n}\cdots\int_{-\infty}^{\lambda_1}p(t_1,\cdots,t_n)\mathrm{d}t_1\cdots\mathrm{d}t_n$$

对所有实值 $\lambda_1,\cdots,\lambda_n$ 成立.

如果 X 和 Y 分别具有均值 m_X 和 m_Y, 它们的协方差 (σ_{XY}) 定义为乘积矩

$$\sigma_{XY} = E[(X - m_X)(Y - m_Y)].$$

如果 X_1 和 X_2 是分别具有分布函数 F_1 和 F_2 的独立随机变量, 则和 $X = X_1 + X_2$ 的分布函数 F 是 F_1 和 F_2 的**卷积**

$$F(x) = \int F_1(x-y)\mathrm{d}F_2(y) = \int F_2(x-y)\mathrm{d}F_1(y).$$

特别当 X_1 和 X_2 具有概率密度 p_1 和 p_2 时, 和 $X = X_1 + X_2$ 的密度函数 p 是密度 p_1 和 p_2 的卷积

4

$$p(x) = \int p_1(x-y)p_2(y)\mathrm{d}y = \int p_2(x-y)p_1(y)\mathrm{d}y.$$

B. 条件分布和条件期望

已知事件 B 的条件下事件 A 的条件概率 $\Pr\{A|B\}$ 定义如下:

$$\Pr\{A|B\} = \frac{\Pr\{AB\}}{\Pr\{B\}}, \quad 若 \Pr\{B\} > 0;$$

当 $\Pr\{B\} = 0$ 时没有定义或指定一个任意值. 设 X 和 Y 是仅取可列多个不同值的随机变量, 例如 $1, 2, \cdots$, **已知 $Y = y$ 时 X 的条件分布函数** $F_{X|Y}(\cdot|y)$ 定义为

$$F_{X|Y}(x|y) = \frac{\Pr\{X \leqslant x, Y = y\}}{\Pr\{Y = y\}}, \quad 若 \Pr\{Y = y\} > 0;$$

当 $\Pr\{Y = y\} = 0$ 时, 规定为任意离散型分布函数. 后面这个规定与以后引用条件分布函数的计算相一致.

若 X 和 Y 具有联合概率密度函数 $P_{XY}(x, y)$, 则已知 $Y = y$ 时 X 的条件分布为

$$F_{X|Y}(x|y) = \frac{\displaystyle\int_{\xi \leqslant x} p_{XY}(\xi, y)\mathrm{d}\xi}{p_Y(y)},$$

这里 $p_Y(y) > 0$; 当 $p_Y(y) = 0$ 时, 可指定为任意值.

注意, $F_{X|Y}$ 满足

(C.P.1) 对每个固定的 $y, F_{X|Y}(x|y)$ 是 x 的分布函数;

(C.P.2) 对每个固定的 $x, F_{X|Y}(x|y)$ 是 y 的函数;

(C.P.3) 对任意 x, y,

$$\Pr\{X \leqslant x, Y \leqslant y\} = \int_{\eta \leqslant y} F_{X|Y}(x|\eta)\mathrm{d}F_Y(\eta),$$

其中, $F_Y(\eta)=\Pr\{Y\leqslant\eta\}$ 是 Y 的边缘分布. 如前面说明的那样, 在本书中读者只须对离散型和连续型随机变量处理 (C.P.3) 中的积分. 即当 Y 有概率密度函数 $p_Y(y)$ 时, (C.P.3) 变为

$$\Pr\{X\leqslant x,Y\leqslant y\}=\int_{\eta\leqslant y}F_{X|Y}(x|\eta)p_Y(\eta)\mathrm{d}\eta;$$

5

而当 Y 是离散型 r.v. 时, (C.P.3) 变为

$$\Pr\{X\leqslant x,Y\leqslant y\}=\sum_{\eta\leqslant y}F_{X|Y}(x|\eta)\Pr\{Y=\eta\}.$$

上述三个性质是条件分布的本质特征. 事实上, 由 (C.P.3) 我们得到

$$\begin{aligned}\Pr\{X\leqslant x,Y=y\}&=\Pr\{X\leqslant x,Y\leqslant y\}-\Pr\{X\leqslant x,Y<y\}\\&=\sum_{\eta\leqslant y}F_{X|Y}(x|\eta)\Pr\{Y=\eta\}-\sum_{\eta<y}F_{X|Y}(x|\eta)\Pr\{Y=\eta\}\\&=F_{X|Y}(x|y)\Pr\{Y=y\},\end{aligned}$$

当 $\Pr\{Y=y\}>0$ 时上式蕴涵着条件分布的定义 $F_{X|Y}(x|y)=\dfrac{\Pr\{X\leqslant x,Y=y\}}{\Pr\{Y=y\}}$.

在较高级的著作中, (C.P.1~3) 被作为条件分布定义的基础. 由此可以规定任意两个随机变量 X 和 Y 的条件分布, 进而规定任意两个随机向量 $X=(X_1,\cdots,X_n)$ 和 $Y=(Y_1,\cdots,Y_n)$ 的条件分布.[1]

当 $y=\infty$ 时, 应用 (C.P.3) 即导出**全概率公式**

$$\Pr\{X\leqslant x\}=\Pr\{X\leqslant x,Y\leqslant\infty\}=\int_{-\infty}^{+\infty}F_{X|Y}(x|y)\mathrm{d}F_Y(y),$$

这是概率论中最基本的公式之一. 当 Y 是离散型时, 上式变为

$$\Pr\{X\leqslant x\}=\sum_{y}\Pr\{X\leqslant x|Y=y\}\Pr\{Y=y\},$$

当 Y 有概率密度函数 $p_Y(y)$ 时, 我们有

$$\Pr\{X\leqslant x\}=\int_{-\infty}^{+\infty}\Pr\{X\leqslant x|Y=y\}p_Y(y)\mathrm{d}y.$$

6

当 X 和 Y 具有联合概率密度函数 $p_{XY}(x,y)$ 时, 我们可以定义已知 $Y=y$ 时 X 的条件密度函数为

$$p_{X|Y}(x|y)=\frac{\mathrm{d}}{\mathrm{d}x}F_{X|Y}(x|y)=\frac{p_{XY}(x,y)}{p_Y(y)},$$

1. 进一步地阐述, 包括关于条件期望严格和直观的讨论, 可参阅 6.7 节. 这些概念对发展现代鞅理论起着重要作用.

这里假定 $p_Y(y)>0$; 当 $p_Y(y)=0$ 时可指定为任意概率密度函数.

设 g 是一实函数, 且 $g(X)$ 的期望是有限的. **已知 $Y=y$ 时 $g(X)$的条件期望**可表达为

$$E[g(X)|Y=y]=\int_x g(x)\mathrm{d}F_{X|Y}(x|y).$$

当 X,Y 具有联合概率密度函数时, $E[g(X)|Y=y]$ 可通过下式计算

$$\begin{aligned}E[g(X)|Y=y]&=\int g(x)p_{X|Y}(x|y)\mathrm{d}x\\&=\frac{\int g(x)p_{XY}(x,y)\mathrm{d}x}{p_Y(y)},\quad 若\ p_Y(y)>0;\end{aligned}\tag{1.5}$$

若 X 和 Y 是离散型随机变量, X 可能取值 $x_1,x_2,\cdots$, 则上式归结为

$$\begin{aligned}E[g(X)|Y=y]&=\sum_{i=1}^{\infty}g(x_i)\Pr\{X=x_i|Y=y\}\\&=\frac{\sum_{i=1}^{\infty}g(x_i)\Pr\{X=x_i,Y=y\}}{\Pr\{Y=y\}},\quad 若\ \Pr\{Y=y\}>0.\end{aligned}$$

与 (C.P.1~3) 类似, 易知在给定 $Y=y$ 时 $g(X)$ 的条件期望满足

(C.E.1) 对于每个函数 g, 若 $E[|g(X)|]<+\infty$, 则 $E[g(X)|Y=y]$ 是 y 的函数;

(C.E.2) 对于任意有界函数 h, 我们有

$$E[g(X)h(Y)]=\int E[g(X)|Y=y]h(y)\mathrm{d}F_Y(y),$$

其中 F_Y 是 Y 的边缘分布函数.

下面我们证实在连续型情况下公式 (C.E.2) 成立.

7

假设使 $p_Y(y)>0$ 的 y 值的集合是一个区间 $(a,b),-\infty\leqslant a<b\leqslant+\infty$. 利用 $F_Y(y)$ 具有概率密度函数, 并用式 (1.5) 代入, 得

$$\begin{aligned}\int E[g(X)|Y=y]h(y)\mathrm{d}F_Y(y)&=\int_a^b E[g(X)|Y=y]h(y)p_Y(y)\mathrm{d}y\\&=\int_a^b\left(\int_{-\infty}^{+\infty}g(x)p_{X|Y}(x|y)\mathrm{d}x\right)h(y)p_Y(y)\mathrm{d}y\\&=\int_a^b\left(\int_{-\infty}^{+\infty}g(x)\frac{p_{XY}(x,y)}{p_Y(y)}\mathrm{d}x\right)h(y)p_Y(y)\mathrm{d}y\\&=\int_a^b\int_{-\infty}^{+\infty}g(x)h(y)p_{XY}(x,y)\mathrm{d}x\mathrm{d}y\\&=E[g(X)h(Y)].\end{aligned}$$

最后一个等号用到假定：仅当 $a < y < b$ 时 $p_{XY}(x,y) > 0$.

在 (C.E.2) 中，若 $h(y) \equiv 1$，即导出对于期望的**全概率公式**

$$E[g(X)] = \int E[g(X)|Y = y]\mathrm{d}F_Y(y).$$

当 Y 离散型时上式变为

$$E[g(X)] = \sum_{i=1}^{\infty} E[g(X)|Y = y_i]\Pr\{Y = y_i\},$$

当 Y 具有概率密度函数 p_Y 时变为

$$E[g(X)] = \int E[g(X)|Y = y]p_Y(y)\mathrm{d}y.$$

由于已知 $Y = y$ 时 $g(X)$ 的条件期望是关于条件分布 $F_{X|Y}$ 的期望，因此条件期望的性质类似于期望. 特别，若 a_1 和 a_2 是固定常数，g_1 和 g_2 为已知函数且满足 $E[|g_i(X)|] < \infty, i = 1, 2$，则

$$\begin{aligned}&E[a_1g_1(X) + a_2g_2(X)|Y = y]\\&= a_1E[g_1(X)|Y = y] + a_2E[g_2(X)|Y = y].\end{aligned}$$

由 (C.E.1) 知，$E[g(X)|Y = y]$ 是实变量 y 的函数，若将这个函数视为随机变量 Y 的函数，我们便得到一个随机变量，记为 $E[g(X)|Y]$. 基本性质 (C.E.2) 可改述为，对任何 y 的有界函数 h，有

$$E[g(X)h(Y)] = E\{E[g(X)|Y]h(Y)\}.$$

8

当 $h(y) = 1$ 时，我们得到如下形式的全概率公式

$$E[g(X)] = E\{E[g(X)|Y]\}.$$

作为小结，下面我们列出条件期望的上述性质和其他性质. 这里带下标和不带下标的 X 和 Y 均表示随机变量，c 是实数，g 是满足 $E[|g(X)|] < \infty$ 的实函数，f 是有界函数，h 是满足 $E[|h(X, Y)|] < \infty$ 的二元实变量函数.

$$\begin{aligned}&E[a_1g(X_1) + a_2g(X_2)|Y]\\&= a_1E[g(X_1)|Y] + a_2E[g(X_2)|Y],\end{aligned} \tag{1.6}$$

$$g \geqslant 0 \text{ 意味着 } E[g(X)|Y] \geqslant 0, \tag{1.7}$$

$$E[h(X,Y)|Y=y]=E[h(X,y)|Y=y], \tag{1.8}$$

$$\text{若 } X \text{ 和 } Y \text{ 独立, 则 } E[g(X)|Y]=E[g(X)], \tag{1.9}$$

$$E[g(X)f(Y)|Y]=f(Y)E[g(X)|Y], \tag{1.10}$$

$$E[g(X)f(Y)]=E\{E[g(X)|Y]f(Y)\}, \tag{1.11}$$

作为 (1.6), (1.10) 和 (1.11) 的推论, 令 $g\equiv 1$ 或 $f\equiv 1$, 我们得到

$$E[c|Y]=c, \tag{1.12}$$

$$E[f(Y)|Y]=f(Y), \tag{1.13}$$

且

$$E[g(X)]=E\{E[g(X)|Y]\}. \tag{1.14}$$

C. 随机变量的无限族

在处理随机变量的无限族问题时, 直接推广前面的定义遇到了实质性的困难. 我们必须稍微修改一下前面的方法.

给定可列随机变量族 $X_1, X_2, \cdots$, 它们的统计性质确定如下, 对每个整数 $n\geqslant 1$ 和任意的 n 个不同的正整数 $i_1, i_2, \cdots, i_n$, 随机变量 $X_{i_1}, X_{i_2}, \cdots, X_{i_n}$ 的联合分布函数为 $F_{X_{i_1},X_{i_2},\cdots,X_{i_n}}$ 当然, 对分布函数族 $F_{X_{i_1},X_{i_2},\cdots,X_{i_n}}$ 应当附加某一相容性条件, 即

9

$$\begin{aligned}&F_{X_{i_1},\cdots,X_{i_{j-1}},X_{i_{j+1}},\cdots,X_{i_n}}(\lambda_1,\cdots,\lambda_{j-1},\lambda_{j+1},\cdots,\lambda_n)\\&=\lim_{\lambda_j\to\infty}F_{X_{i_1},\cdots,X_{i_n}}(\lambda_1,\cdots,\lambda_{j-1},\lambda_j,\lambda_{j+1},\cdots,\lambda_n),\end{aligned}$$

以及, 任意调换分布函数 $F_{X_{i_1},X_{i_2},\cdots,X_{i_n}}$ 中的两个下标 i_ν 和 i_μ 及其对应的变量 λ_ν 和 λ_μ, 所得到的分布函数不变. 这意味着随机变量列 $X_1, X_2, \cdots$ 与所附下标方式无关.

联合分布函数族 $\{F_{X_{i_1},X_{i_2},\cdots,X_{i_n}}\}$ 称为 $\{X_n\}_{n=1}^{\infty}$ 的**有限维分布函数族**, 原则上说, 有关随机变量族 $\{X_n\}_{n=1}^{\infty}$ 的所有重要概率均可由有限维分布函数族来计算.

D. 特征函数

与 r.v. X 的分布函数相联系的重要函数是它的特征函数 $\phi(t)$(简记为 c.f.), 这里 t 是实变量 $-\infty<t<+\infty$. 我们把它写成如下形式

$$\begin{aligned}\phi(t)&=\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}t\lambda}\mathrm{d}F(\lambda),\quad \mathrm{i}=\sqrt{-1}\\&=E[\mathrm{e}^{\mathrm{i}tX}].\end{aligned} \tag{1.15}$$

读者不妨先形式上理解 (1.15). 如果 F 有概率密度函数 p, 特征函数为

$$\phi(t)=\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}t\lambda}p(\lambda)\mathrm{d}\lambda.$$

若 F 是离散型随机变量 X 的分布, X 可能取值为 $\{\lambda_k\}_{k=0}^{\infty}$, 且 $\Pr\{X=\lambda_k\}=a_k(k=0,1,\cdots)$, 则式 (1.15) 化为级数

$$\phi(t)=\sum_{k=0}^{\infty}\mathrm{e}^{\mathrm{i}t\lambda_k}a_k.$$

特征函数的重要意义大体可以从下述三个方面看出:

(a) 分布函数和特征函数之间的关系是一一对应的. 这样, 知道了特征函数等于知道了分布函数. 用特征函数表达分布函数的方程式就是著名的勒维 (Levy) 变换公式. 由于我们不用它, 对这个问题的讨论请读者阅读有关参考书.

(b) 如果 $X_1,\cdots,X_n$ 是 n 个相互独立的随机变量, 它们和的特征函数便是它们的特征函数的乘积. 这个简单的结果使特征函数在处理有关独立随机变量和问题时非常方便. 10

(c) 若随机变量的矩有限时, 它们可由特征函数的微商得到, 其关系式是

$$E[X^k]=\frac{1}{\mathrm{i}^k}\phi^{(k)}(0),$$

其中, $\mathrm{i}=\sqrt{-1},\phi^{(k)}(t)=\mathrm{d}^k\phi(t)/\mathrm{d}t^k$ 是 c.f. $\phi(t)$ 的第 k 阶导数.

在各种极限过程中, 分布函数和它们的特征函数之间仍然保持这种一对一关系. 事实上, 如果 $F,F_1,F_2,\cdots$ 是分布函数, 且对 F 的每个连续点 λ, $\lim\limits_{n\to\infty}F_n(\lambda)=F(\lambda)$, $\phi_n(t)$ 是 F_n 的特征函数, 则在每个有限区间上一致成立

$$\phi_n(t)=\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}t\lambda}\mathrm{d}F_n(\lambda)\longrightarrow\phi(t)=\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}t\lambda}\mathrm{d}F(\lambda);$$

反之, 如果 $\phi_1,\phi_2,\cdots$ 是分布函数 $F_1,F_2,\cdots$ 的特征函数, 且对每个 t 有 $\lim\limits_{n\to\infty}\phi_n(t)=\phi(t),\phi(t)$ 在 $t=0$ 是连续的, 则 $\phi(t)$ 是分布函数 F 的特征函数, 并且 $\lim\limits_{n\to\infty}F_n(\lambda)=F(\lambda)$ 对 F 的每一连续点成立. 这个结果是著名的 Levy 收敛准则.

E. 母函数和拉普拉斯变换

对于仅取非负整数值的随机变量, 与特征函数相应的一个函数是母函数, 其定义为

$$g(s)=\sum_{k=0}^{\infty}p_ks^k=E[s^X],$$

其中

$$p_k=\Pr\{X=k\}.$$

由假设 $p_k\geqslant 0$ 且 $\sum\limits_{k=0}^{\infty}p_k=1,g(s)$ 至少对于 $|s|\leqslant 1$(s 是复变量) 是有定义的, 且对于 $|s|<1$ 是无限次可微的. 非负整数值随机变量的母函数可经过变量替换 $s=\mathrm{e}^{\mathrm{i}t}$ 变成 X 的特征函数 ϕ: 11

$$\phi(t) = E[\mathrm{e}^{\mathrm{i}tX}] = E[(\mathrm{e}^{\mathrm{i}t})^X] = g(\mathrm{e}^{\mathrm{i}t}).$$

这样, 母函数保持了特征函数的三个基本性质:

(a) 母函数唯一确定其分布函数;

(b) 独立非负整数值随机变量和的母函数是它们母函数的乘积;

(c) 各阶矩可通过逐次微分得到, 阶乘矩可由下式得到

$$E[X(X-1)\cdots(X-k)] = g^{(k+1)}(1),$$

这里 $g^{(k)}(s) = \mathrm{d}^k g(s)/\mathrm{d}s^k$ 是 g 的第 k 阶导数, 因此

$$E[X] = g^{(1)}(1),$$

且

$$E[X^2] = g^{(2)}(1) + g^{(1)}(1).$$

下面我们举例说明如何使用母函数来处理独立随机变量和的问题. 设 $N, X_1, X_2, \cdots$ 是独立非负整数值随机变量. 现要确定和 $R = X_1 + X_2 + \cdots + X_N$ 的母函数 $g_R(s)$. 这是一个具有随机项数的随机变量和.

令 $g_N(s)$ 是 N 的母函数, 并假定 X_i 同分布, 具有相同的母函数 $g(s)$. 利用 (1.14) 和 (1.9),

$$\begin{aligned} g_R(s) &= E[s^R] \\ &= E[s^{X_1+\cdots+X_N}] \\ &= E\{E[s^{X_1+\cdots+X_N}|N]\} \\ &= \sum_{n=0}^{\infty} E[s^{X_1+\cdots+X_n}|N=n]\Pr\{N=n\} \\ &= \sum_{n=0}^{\infty} E[s^{X_1+\cdots+X_n}]\Pr\{N=n\} \end{aligned}$$

(因为 N 和 X_i 是独立的)

$$\begin{aligned} &= \sum_{n=0}^{\infty} g^n(s)\Pr\{N=n\} \\ &= E[g^N(s)] \\ &= g_N[g(s)]. \end{aligned}$$

12

综上,

$$g_R(s) = g_N[g(s)].$$

利用复合函数微分链式法则, 有

$$g_R'(s) = g_N'[g(s)] \cdot g'(s),$$

令 $s = 1$, 我们推得

$$E[R] = E[N] \cdot E[X].$$

用类似方式计算 R 的方差 σ_R^2, 我们得到

$$\sigma_R^2 = E[X]^2\sigma_N^2 + E[N]\sigma_X^2,$$

这里 σ_N^2 和 σ_X^2 分别是 N 和 X 的方差 (见初等问题 4).

把上面结果进一步推广. 设 $X_1, X_2, \cdots$ 是任意独立同分布随机变量 (即未必取整数值) 而 N 仍旧如上述. 则

$$\phi_R(t) = g_N(\phi(t)),$$

这里 ϕ_R 和 g_N 分别是 $R = X_1 + \cdots + X_N$ 和 N 的特征函数及母函数, 而 ϕ 是 X_i 共同的特征函数.

在考虑非负随机变量时, 利用分布函数的拉普拉斯变换代替特征函数显得更自然些. 如果分布函数 F_X 有密度函数 p_X, 拉普拉斯变换定义为

$$\psi_X(s) = \int_0^\infty \mathrm{e}^{-sx} p_X(x)\mathrm{d}x.$$

这个积分对于复变量 s 是存在的, 其中 $s = \sigma + \mathrm{i}t, \sigma$ 和 t 是实的, $\sigma \geqslant 0$. 当 s 是纯虚数时, 即 $s = \mathrm{i}t, \psi_X(s)$ 就是特征函数 $\phi_X(-t)$. 对于离散型非负随机变量, 拉普拉斯变换定义为

$$\psi_X(s) = \sum_{n=0}^\infty \mathrm{e}^{-s\lambda_n} \Pr\{X = \lambda_n\}.$$

与特征函数类似, 如果 $X_1, X_2, \cdots, X_n$ 是相互独立的非负随机变量组, 则

$$\psi_{X_1+\cdots+X_n}(s) = \prod_{k=1}^n \psi_{X_k}(s).$$

对于一般的分布函数, 其拉普拉斯变换定义为

$$\psi_X(s) = \int_0^\infty \mathrm{e}^{-s\xi} \mathrm{d}F_X(\xi).$$

13

与特征函数类似, 拉普拉斯变换唯一地确定了分布函数.

F. 分布函数的例子

若干常用分布函数的某些基本性质见表 1 和表 2.

表 1 某些常见的连续型概率分布

连续型分布	密度函数 $p(x)$	参数范围	特征函数 $\phi(t)$	均值	方差
正态分布	$\frac{1}{\sqrt{2\pi}\sigma}\exp-\frac{(x-m)^2}{2\sigma^2}$, $-\infty < x < \infty$	m 实数, $\sigma > 0$	$\exp\left[-\frac{\sigma^2 t^2}{2} + \mathrm{i}mt\right]$	m	σ^2
指数分布	$\lambda e^{-\lambda x}, x > 0$	$\lambda > 0$	$\frac{\lambda}{\lambda - \mathrm{i}t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Γ 分布	$\frac{\lambda}{\Gamma(\alpha)}(\lambda x)^{\alpha-1}e^{-\lambda x}$, $x > 0$	$\lambda > 0$, $\alpha > 0$	$\frac{\lambda^\alpha}{(\lambda - \mathrm{i}t)^\alpha}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
均匀分布	$\frac{1}{b-a}, a < x < b$	$a < b$	$\frac{e^{\mathrm{i}ub} - e^{\mathrm{i}ua}}{\mathrm{i}u(b-a)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Beta 分布	$\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}x^{p-1}(1-x)^{q-1}$, $0 < x < 1$	$p > 0$, $q > 0$	$\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\int_0^1 e^{\mathrm{i}tx}x^{p-1}(1-x)^{q-1}\mathrm{d}x$	$\frac{p}{p+q}$	$\frac{qp}{(p+q)^2(p+q+1)}$

注释： 当 $\alpha = 1$ 时, Γ 分布是指数分布, 其参数 λ 视为标度因子. 当参数 $p = q = 1$ 时, Beta 分布是 (0, 1) 上的均匀分布, 有时简称为 "均匀 (0,1)".

表 2 某些常见的离散型概率分布

离散型分布	概率分布列	参数的可能取值	母函数	均值	方差
泊松分布	$\dfrac{e^{-\lambda}\lambda^n}{n!}, n=0,1,2,\cdots$	$\lambda>0$	$e^{-\lambda+\lambda s}$	λ	λ
二项分布	$\binom{N}{n}p^n q^{N-n}, n=0,1,\cdots,N$	$N=1,2,\cdots$ $0<p<1$ $q=1-p$	$(1-p+ps)^N$	Np	Npq
负二项 (巴斯卡) 分布	$\binom{\alpha+n-1}{n}p^\alpha q^n, n=0,1,2,\cdots$	$\alpha>0$ $0<p<1$	$\left(\dfrac{p}{1-qs}\right)^\alpha$	$\dfrac{\alpha q}{p}$	$\dfrac{\alpha q}{p^2}$
几何分布	$p(1-p)^n, n=0,1,2,\cdots$	$0<p<1$	$\dfrac{p}{1-qs}$	$\dfrac{q}{p}$	$\dfrac{q}{p^2}$

下面给出两个重要的多元分布函数:

(a) *多元正态分布*

设 $\sigma_1,\sigma_2,m_1,m_2$ 和 ρ 是实常数, 并假定 $\sigma_i>0, i=1,2, 0\leqslant|\rho|<1$. 记

$$Q(x_1,x_2)=\frac{1}{1-\rho^2}\left\{\left(\frac{x_1-m_1}{\sigma_1}\right)^2-2\rho\left(\frac{x_1-m_1}{\sigma_1}\right)\left(\frac{x_2-m_2}{\sigma_2}\right)+\left(\frac{x_2-m_2}{\sigma_2}\right)^2\right\},$$

如果 X_1 和 X_2 是随机变量, 联合分布为

$$\Pr\{X_1\leqslant a, X_2\leqslant b\}=\int_{-\infty}^a\int_{-\infty}^b\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2}Q(x_1,x_2)\right\}\mathrm{d}x_1\mathrm{d}x_2$$

则称 X_1,X_2 具有**联合正态分布**. 容易验证, $E[X_i]=m_i (i=1,2)$, X_i 的方差是 σ_i^2, 协方差为

$$E[(X_1-m_1)(X_2-m_2)]=\rho\sigma_1\sigma_2,$$

且 ρ(无量纲变量) 称为**相关系数**. 联合特征函数为

$$\begin{aligned}\phi_{X_1,X_2}(t_1,t_2)&=E[e^{i(t_1X_1+t_2X_2)}]\\&=\exp\left\{i(t_1m_1+t_2m_2)-\frac{1}{2}(t_1^2\sigma_1^2+2\rho t_1\sigma_1t_2\sigma_2+t_2^2\sigma_2^2)\right\}\end{aligned}$$

若 X_1 和 X_2 具有联合正态分布, 则已知 $X_1=x_1$ 时 X_2 的条件分布是具有如下概率密度函数的正态分布

$$p_{X_2|X_1}(x_2|x_1)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x_2-m}{\sigma}\right)^2\right],\quad -\infty<x_2<\infty,$$

其中,

$$m=m_2+\frac{\sigma_2}{\sigma_1}\rho(x_1-m_1)\quad 且\quad \sigma=\sigma_2\sqrt{1-\rho^2}.$$

14
16

设 $\|a_{ij}\|$ 是 $n\times n$ 对称正定矩阵, $\|b_{ij}\|$ 是 $\|a_{ij}\|$ 的逆矩阵, $B=\det\|b_{ij}\|$ 是 $\|b_{ij}\|$ 的行列式. 设 $m_i, i=1,\cdots,n$ 是任意实常数. 称随机变量组 $X_1,\cdots,X_n$ 具有联合正态分布, 如果它们具有如下的联合概率密度函数:

$$p(x_1,\cdots,x_n)=\frac{\sqrt{B}}{(2\pi)^{n/2}}\exp\left\{-\frac{1}{2}Q(x_1,\cdots,x_n)\right\}, -\infty< x_i<\infty,$$

其中,

$$Q(x_1,\cdots,x_n)=\sum_{i,j=1}^{n}(x_i-m_i)b_{ij}(x_j-m_j).$$

其联合特征函数是

$$\begin{aligned}\phi(t_1,\cdots,t_n)&=E\left[\exp\left\{\mathrm{i}\sum_{i=1}^{n}t_iX_i\right\}\right]\\&=\exp\left\{\mathrm{i}\sum_{i=1}^{n}t_im_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}t_ia_{ij}t_j\right\}.\end{aligned}$$

由此可计算

$$E[X_i]=m_i,\quad i=1,\cdots,n$$

和

$$E[(X_i-m_i)(X_j-m_j)]=a_{ij},$$

因而我们称矩阵 $\|a_{ij}\|$ 为**协方差矩阵**.

从特征函数性质易证, 当且仅当 $Y=a_1X_1+\cdots+a_nX_n$ 具有正态分布时, $X_1,\cdots,X_n$ 具有联合正态分布, 这里 $a_1,\cdots,a_n$ 是任意实数.

(b) *多项分布*

多项分布是仅取非负整数值 $0,\cdots,n$ 的 $r(r\geqslant 2)$ 个离散型随机变量的联合分布, 它定义为

$$\Pr\{X_1=k_1,\cdots,X_r=k_r\}=\begin{cases}\dfrac{n!}{k_1!\cdots k_r!}p_1^{k_1}\cdots p_r^{k_r}, & \text{若 } k_1+\cdots+k_r=n\\ 0, & \text{其他},\end{cases}$$

其中 $p_i>0, i=1,\cdots,r$, 且 $\sum_{i=1}^{r}p_i=1$.

它的联合母函数为

$$\begin{aligned}g(s_1,\cdots,s_r)&=E[s_1^{X_1}\cdots s_r^{X_r}]\\&=(p_1s_1+\cdots+p_rs_r)^n.\end{aligned}$$

G. 极限定理

一列实数 $\{a_n\}$ 称为收敛于实数 a, 记为 $\lim\limits_{n\to\infty} a_n = a$, 如果对任意正数 ε, 存在自然数 $N(\varepsilon)$, 使得对所有的 $n > N(\varepsilon)$, $|a_n - a| < \varepsilon$. 有几个途径可以把这个收敛性概念推广于随机变量的情形, 设 $Z, Z_1, Z_2, \cdots$ 是随机变量序列.

(a) 以概率 1 收敛

如果 $\Pr\{\lim\limits_{n\to\infty} Z_n = Z\} = 1$, 则说 Z_n 以概率 1 收敛于 Z. 换句话说, 以概率 1 出现结果: $Z = z, Z_1 = z_1, Z_2 = z_2, \cdots$, 且 $\lim\limits_{n\to\infty} z_n = z$.

(b) 依概率收敛

如果对每个正数 ε, 成立 $\lim\limits_{n\to\infty} \Pr\{|Z_n - Z| > \varepsilon\} = 0$, 或者 $\lim\limits_{n\to\infty} \Pr\{|Z_n - Z| \leqslant \varepsilon\} = 1$, 则说 Z_n 依概率收敛于 Z. 换句话说, 我们可以取足够大的 n, 使得 Z_n 任意接近 Z 的概率可达任意大.

(c) 二次平均收敛 (简称均方收敛)

如果 $\lim\limits_{n\to\infty} E[|Z_n - Z|^2] = 0$ 则说 Z_n 二次平均收敛于 Z(简称 Z_n 均方收敛于 Z). 换句话说, 我们可以取足够大的 n, 在差的平方平均意义下 Z_n 任意逼近 Z.

(d) 依分布收敛

设 $F(t) = \Pr\{Z \leqslant t\}$ 和 $F_k(t) = \Pr\{Z_k \leqslant t\}$, $k = 1, 2, \cdots$, 如果 $\lim\limits_{n\to\infty} F_n(t) = F(t)$ 对所有 F 的连续点 t 成立, 则说 Z_n 依分布收敛于 Z.

可以证明, Z_n 以概率 1 收敛于 Z 可推出 Z_n 依概率收敛于 Z, 后者又可推得 Z_n 依分布收敛于 Z. 这样, 依分布收敛是收敛性中最弱的形式. 事实上, 可以证明每个分布函数族 $\{F_\alpha\}$ 包含一个序列 $\{F_{\alpha_n}\}$ 使得在 F 的连续点 t 上收敛于 F(Helly-Bray 引理), 但这里 F 不一定是正常的分布函数, $F(\infty)$ 可能小于 1. 18

概率论的很多结果是以极限定理的形式出现的, 下面仅叙述其中的几个. (我们不叙述那些最弱条件下的结果.)

设 $X_1, X_2, \cdots$ 是独立同分布随机变量序列, 具有有限均值 m. 令 $S_n = X_1 + \cdots + X_n$, $\overline{X}_n = S_n/n$ 是样本均值.

弱大数定理 $\overline{X}_n$ 依概率收敛于 m. 即, 对任意 $\varepsilon > 0$, 有

$$\lim_{n\to\infty} \Pr\{|\overline{X}_n - m| > \varepsilon\} = 0.$$

强大数定理 $\overline{X}_n$ 以概率 1 收敛于 m. 即,

$$\Pr\{\lim_{n\to\infty} \overline{X}_n = m\} = 1.$$

中心极限定理 假设 X_k 具有有限方差 σ^2, 设

$$Z_n = \frac{S_n - nm}{\sigma\sqrt{n}} = \frac{1}{\sigma}(\overline{X}_n - m)\sqrt{n},$$

记 Z 是零均值、单位方差的正态随机变量, 则 Z_n 依分布收敛于 Z. 即, 对于任意实数 a,

$$\lim_{n\to\infty} \Pr\{Z_n \leqslant a\} = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-u^2/2} \mathrm{d}u.$$

Borel-Cantelli 引理 设 $A_1, A_2, \cdots$ 是独立事件的无限序列. 则事件 $\{A_i, \text{i.o.}\}$ (这里, i.o. 表示无限地经常, 意指有无限多数目的事件 A_i 发生) 可由下式表达

$$\overline{A}_\infty = \{A_i, \text{i.o.}\} = \bigcap_{j=1}^{\infty} \bigcup_{i=j}^{\infty} A_i.$$

Borel-Cantelli 引理指出, $\overline{A}_\infty$ 的概率或是 0 或是 1, 分别依 $\sum\limits_{i=1}^{\infty} \Pr\{A_i\} < \infty$ 或

19 $\sum\limits_{i=1}^{\infty} \Pr\{A_i\} = \infty$ 而定.

H. 不等式

有若干不等式在研究随机过程中起着重要作用. 我们仅叙述其中的两个.

切比雪夫不等式 设 Z 是非负随机变量. 则对任何正数 c

$$\Pr\{Z > c\} \leqslant \frac{1}{c} E[Z]. \tag{1.16}$$

证明 由于 Z 是非负的,

$$\begin{aligned} E[Z] &= \int_0^{\infty} z\mathrm{d}F(z) \geqslant \int_c^{\infty} z\mathrm{d}F(z) \\ &\geqslant c \cdot \int_c^{\infty} \mathrm{d}F(z) = c \cdot \Pr\{Z > c\}. \end{aligned}$$

由此即推得要证的不等式. 如果 X 是具有均值 μ 和方差 σ^2 的随机变量, 我们应用 (1.16) 于 $Z = (X - \mu)^2$, 使得

$$\Pr\{Z > \varepsilon^2\} = \Pr\{|X - \mu| > \varepsilon\} \leqslant \frac{\sigma^2}{\varepsilon^2}.$$

施瓦兹不等式 设 X, Y 是具有有限二阶矩的随机变量, 则

$$(E[XY])^2 \leqslant E[X^2]E[Y^2].$$

证明 对任意实数 λ,

$$0 \leqslant E[(X+\lambda Y)^2] = E[X^2] + 2\lambda E[XY] + \lambda^2 E[Y^2].$$

把上式看作 λ 的二次函数, 那么它至多有一个实数根. 因而此二项式的判别式应当是非正的, 即

$$4(E[XY])^2 \leqslant 4E[X^2]E[Y^2],$$

这就完成了证明.

1.2 随机过程的两个简单例子

本书将对随机过程各方面作导引式的介绍. 随机过程理论主要研究随机变量族 $\{X_t\}$ 的构造, 这里, 参数 t 取遍某个指标集 T. 在不引起混淆的情况下, 有时用 $X(t)$ 代替 X_t.

20

随机过程 $\{X_t, t \in T\}$ 的一个**实现或样本函数**, 是指对每个 $t \in T$, X_t 的可能值的一个指定. 指标集可以是离散时间 $T = \{0, 1, 2, \cdots\}$, 这时 $\{X_t\}$ 可以表示一系列试验的结果, 例如, 抛掷硬币的一系列结果, 智力测验时对象的一系列反应, 或者对人口某种特征的一系列观察, 等等.

X_t 的值可以是一维的, 二维的或者 n 维的, 甚至更加一般化. 当 X_n 表示第 n 次抛掷骰子的结果, 它的可能值集合是 $\{1, 2, 3, 4, 5, 6\}$ 这个过程的一个实现, 譬如说是 5, 1, 3, 2, 2, 4, 1, 6, 3, 6, $\cdots$, 可直观表示如图 1-1, 其中 $t = n$ 的纵坐标是 X_n 的值. 在这个例子中, 随机变量 X_n 是相互独立的. 但一般地说, 随机变量 X_n 并不相互独立.

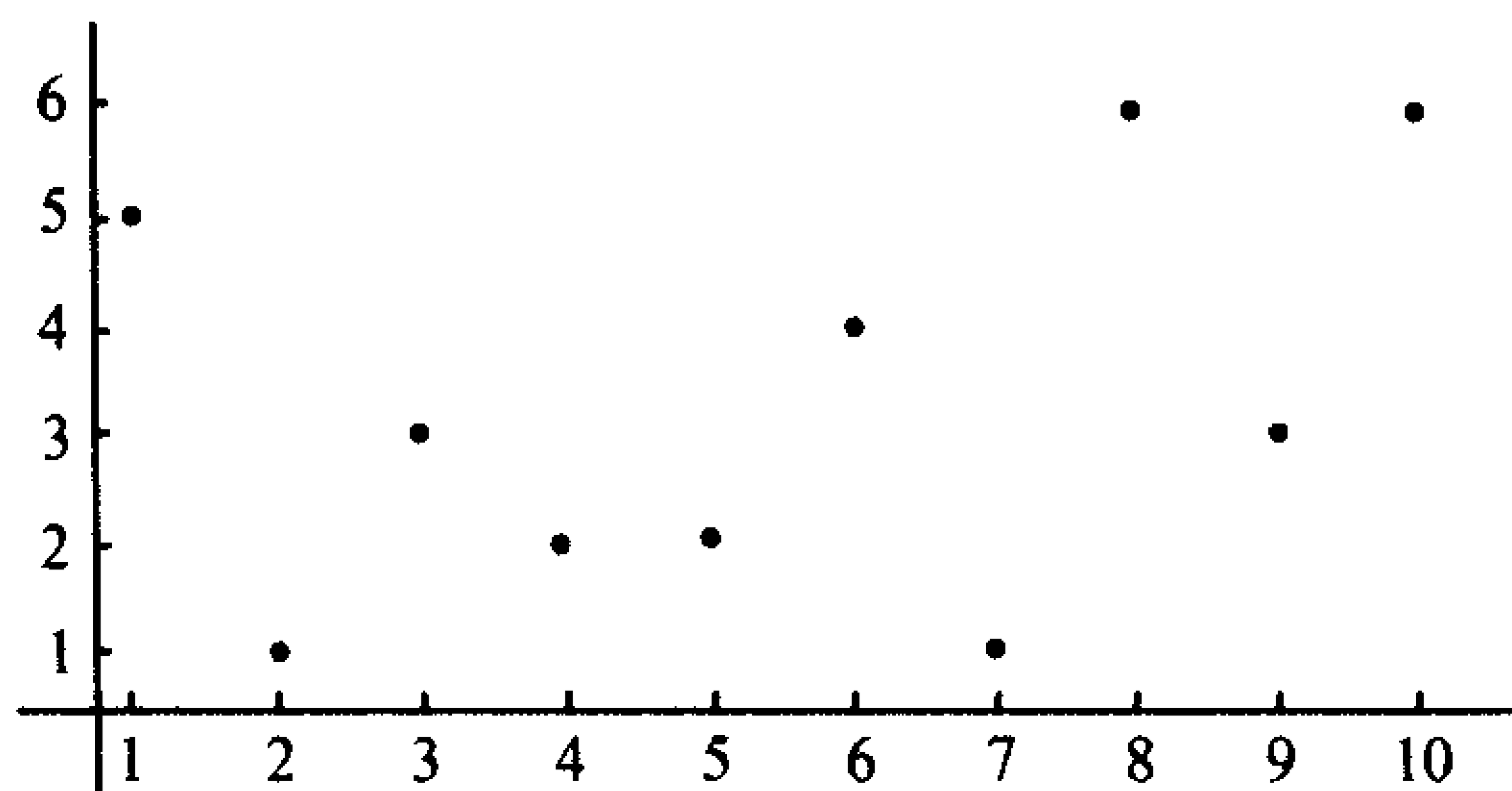


图 1-1

指标集 $T = [0, \infty)$ 的随机过程在应用中尤为重要. 此时 t 通常解释为时间.

我们目前先局限于粗略讨论随机过程及其两个实例的若干基本概念, 本章末将简要介绍不同类型的随机过程, 至于这些例子本身的详尽讨论将在后继有关章节中进行.

例 1 一个非常重要的例子是著名的**布朗运动**. 这个过程具有如下的特征:

(a) 假定 $t_0 < t_1 < \cdots < t_n$, 则增量 $X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \cdots, X_{t_n} - X_{t_{n-1}}$ 是相互独立的随机变量. (具有这种性质的过程称为独立增量过程, 它表示在不重叠的时间段上, X_t 的改变量是相互独立的随机变量.)

21

(b) 当 $t_2 > t_1, X_{t_2} - X_{t_1}$ 的概率分布仅仅与 $t_2 - t_1$ 有关 (而与 t_1 无关).

(c) $\Pr[X_t - X_s \leqslant x] = [2\pi B(t-s)]^{-1/2} \int_{-\infty}^{x} \exp[-u^2/2B(t-s)]\mathrm{d}u$ 其中, $t > s$, B 是正常数.

假设对每一条轨道 $X_0 = 0$. 这时, $E[X_t] = 0, \sigma^2(X_t) = Bt$, 这里 B 是固定的正常数. 可以证明, 如果 $0 < t_1 < t_2 < \cdots < t_n < t$, 在已知 $X_{t_1}, X_{t_2}, \cdots, X_{t_n}$ 时 X_t 的条件概率分布由下式给出 (见第 7 章)

$$\begin{aligned}
&\Pr\{X_t \leqslant x | X_{t_1} = x_1, \cdots, X_{t_n} = x_n\} \\
&\quad = [2\pi B(t-t_n)]^{-\frac{1}{2}} \int_{-\infty}^{x-x_n} \exp[-u^2/2B(t-t_n)]\mathrm{d}u.
\end{aligned}$$

这个过程的历史始于 1827 年布朗对浸于液体中的小粒子不停地无规则运动的观察. 1905 年爱因斯坦假定所观察的粒子是受到周围介质分子不断地碰撞而解释了这种运动. 由爱因斯坦导出的解析结果后来又被许多物理学家及数学家进一步验证和充实.

设 X_t 表示布朗运动粒子在时刻 t 的位移 (从某个出发点开始, 沿着某固定轴). 在时间区间 (s,t) 上位移 $X_t - X_s$ 可以看作是大量小位移之和, 因此利用中心极限定理, 断言 $X_t - X_s$ 呈现正态分布看来是合理的. 如果假定介质是均匀的, 则我们可以认为 $X_t - X_s$ 和 $X_{t+h} - X_{s+h}$ 的分布对任意 $h > 0$ 都是相同的. 最后, 由直观易知位移 $X_t - X_s$ 仅仅取决于时间间隔 $t - s$, 而与我们开始观察的时刻无关.

布朗运动 (也称为 Wiener 过程) 是研究许多其他类型随机过程的基础. 第 7 章将详细讨论一维布朗运动.

例 2 连续时间 ($T = [0,\infty)$) 随机过程的另一重要例子是**泊松过程**. 其样本函数 X_t 表示从 0 到 t 这段时间内所研究的事件出现在次数, 于是, 每个样本函数 X_t 是不减的阶梯函数.

22

图 1-2 表示这样的一个样本函数, 事件首先出现在时刻 t_1, 然后在时刻 t_2, t_3, t_4 等等. 显然事件出现的总数仅以单位跳跃式增加, 并且 $X_0 = 0$. 这种过程的实例很多, 如, 一个放射性物质在衰变中所散发的 X 射线的数目; 发生于一个地区的电话呼唤次数; 某个十字路口事故出现的次数; 在一页内打字差错出现的次数; 一台机器损坏的次数; 到达服务台的顾客的数目. 把上述这些例子看成泊松过程的依据是所谓的稀有事件律. 若我们有试验次数很多的贝努里试验, 每次试验成功的概率很小,

而期望成功的次数是一常数, 在这种条件下, 由大家熟悉的概率论定理知道, 事件实际出现次数近似遵循泊松规律. 在放射性物质衰变中, 如果观察的时间相对于放射性物质半衰期是非常短的, 采用泊松逼近非常理想.

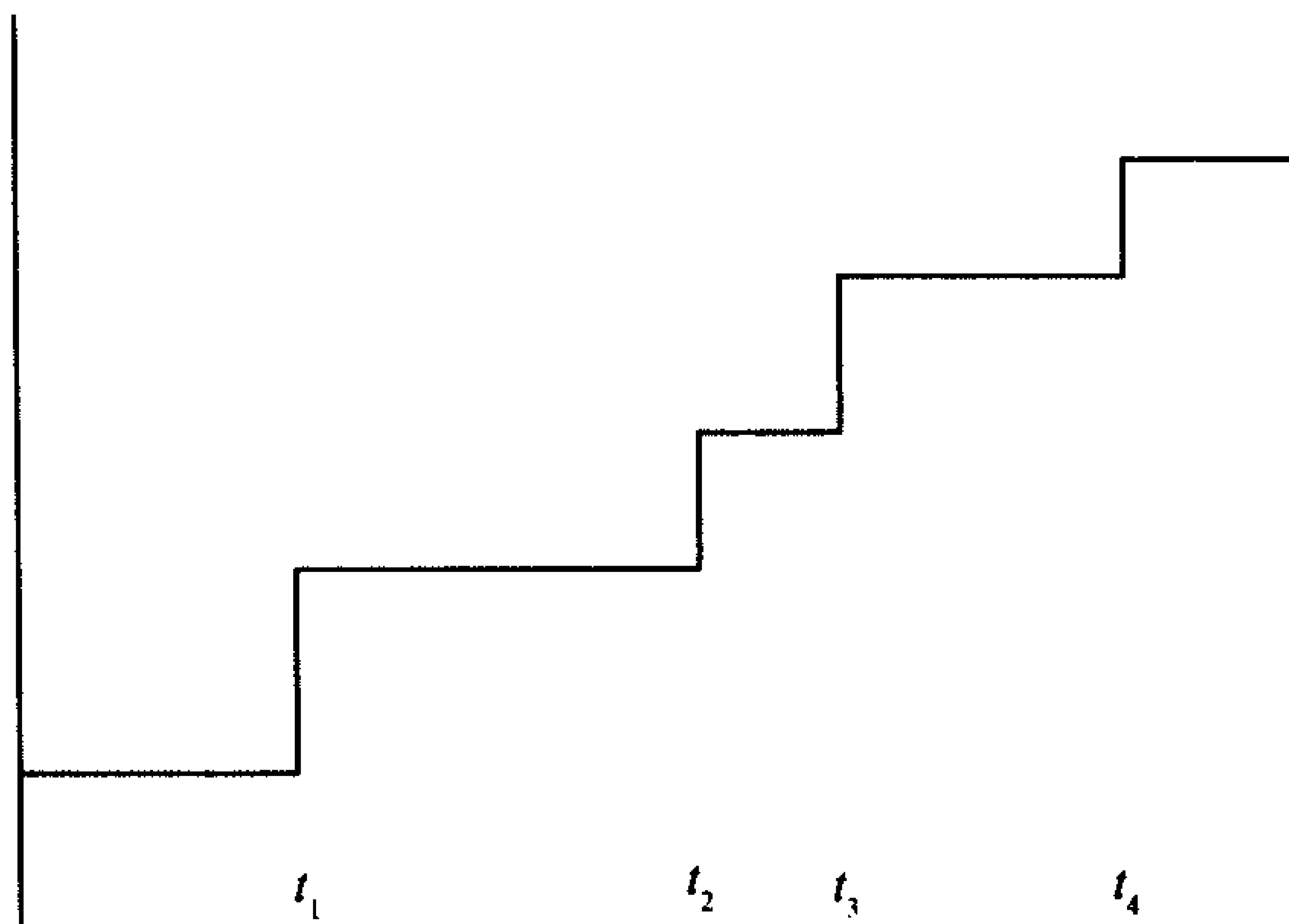


图 1-2

我们假定发生在两个不相交时间区间的事件次数相互独立 (见 (a)). 类似于 (b), 我们也假定随机变量 $X_{t_0+t} - X_{t_0}$ 的概率分布只取决于 t 而与 t_0 无关. 与上面直观描述相一致, 我们还进一步给出如下假定:

I. 在长度为 h 的时间区间中, 事件至少发生一次的概率为

$$p(h) = ah + o(h), \quad h \to 0, a > 0$$

23

$[g(t) = o(t), t \to 0$, 表示 $\lim\limits_{t\to 0} g(t)/t = 0]$.

II. 在长度为 h 的时间区间中, 事件发生两次或两次以上的概率为 $o(h)$.

假定 II 相当于排除了同时发生两次或更多次事件的可能性. 在上列例子中此要求通常可以满足.

设 $P_m(t)$ 表示在 t 时间内事件恰好发生 m 次的概率, 即

$$P_m(t) = \Pr\{X_t = m\}, \quad m = 0, 1, 2, \cdots.$$

假定 II 可表示成下面形式

$$\sum_{m=2}^{\infty} P_m(h) = o(h),$$

而

$$p(h) = P_1(h) + P_2(h) + \cdots.$$

由于独立性的假定,

$$P_0(t+h)=P_0(t)P_0(h)=P_0(t)(1-p(h)),$$

因此

$$\frac{P_0(t+h)-P_0(t)}{h}=-P_0(t)\frac{p(h)}{h}.$$

但由于假定 I , 我们有 $p(h)/h\to a$. 由此推得在 $(0,t)$ 内事件不发生的概率 $P_0(t)$ 满足微分方程

$$P_0'(t)=-aP_0(t),$$

它的解是众所周知的, 即 $P_0(t)=c\mathrm{e}^{-at}$. 常数 c 由初始条件 $P_0(0)=1$ 确定, 推得 $c=1$. 于是 $P_0(t)=\mathrm{e}^{-at}$. 下面我们对任意的 m 计算 $P_m(t)$. 显然有

$$P_m(t+h)=P_m(t)P_0(h)+P_{m-1}(t)P_1(h)+\sum_{i=2}^{m}P_{m-i}(t)P_i(h). \tag{2.1}$$

由定义, $P_0(t)=1-p(h)$. 因 $P_k(t)\leqslant 1$, 由条件 II 推得

24

$$P_1(h)=p(h)+o(h)\quad 及$$

$$\sum_{i=2}^{m}P_{m-i}(t)P_i(h)\leqslant\sum_{i=2}^{m}P_i(h)=o(h). \tag{2.2}$$

借助于 (2.2), 我们重新整理 (2.1) 成为

$$\begin{aligned}P_m(t+h)-P_m(t)&=P_m(t)[P_0(h)-1]+P_{m-1}(t)P_1(h)+\sum_{i=2}^{m}P_{m-i}(t)P_i(h)\\&=-P_m(t)p(h)+P_{m-1}(t)P_1(h)+\sum_{i=2}^{m}P_{m-i}(t)P_i(h)\\&=-aP_m(t)h+aP_{m-1}(t)h+o(h).\end{aligned}$$

因此,

$$\frac{P_m(t+h)-P_m(t)}{h}\longrightarrow -aP_m(t)+aP_{m-1}(t),\quad (h\to 0),$$

从形式上我们得到

$$P_m'(t)=-aP_m(t)+aP_{m-1}(t),\quad m=1,2,\cdots, \tag{2.3}$$

其初始条件为

$$P_m(0)=0,\quad m=1,2,\cdots.$$

为了解方程 (2.3), 我们引入函数

$$Q_m(t) = P_m(t)\mathrm{e}^{at}, \quad m = 0, 1, 2, \cdots.$$

把此式代入 (2.3), 得

$$Q'_m(t) = aQ_{m-1}(t), \quad m = 1, 2, \cdots, \tag{2.4}$$

其中 $Q_0(t) \equiv 1$, 并且初始条件为 $Q_m(0) = 0, m = 1, 2, \cdots$. 应用递归法解 (2.4), 得

$$\begin{aligned}
&Q'_1(t) = a \text{ 或 } Q_1(t) = at + c, && \text{所以 } Q_1(t) = at \\
&Q_2(t) = \frac{a^2t^2}{2} + c, && \text{所以 } Q_2(t) = \frac{a^2t^2}{2!} \\
&\quad\vdots && \quad\vdots \\
& && Q_m(t) = \frac{a^mt^m}{m!},
\end{aligned}$$

故

$$P_m(t) = \frac{a^mt^m}{m!}\mathrm{e}^{-at}.$$

换句话说, 对每个 t, X_t 遵循参数为 at 的泊松分布. 特别地, 在时间 t 事件出现次数的均值为 at.

25

泊松过程中的时间参数还经常被适当的空间位置参数形式所代替. 下面这个例子说明了这方面问题处理的方法. 考虑分布在空间 E(E 是维数 $d \geqslant 1$ 的欧几里得空间) 上点的排列. 令 N_R 表示 E 的某区域 R 内所含的点数 (有限或无限). 我们假设 N_R 是随机变量, 其中 R 取遍 E 的所有可能的子集. 随机变量的集合 $\{N_R\}$, 称为齐次泊松过程, 如果满足下列条件:

(i) 在互不重叠的区域内, 点的数目是相互独立的随机变量.

(ii) 对于任意有限 "体积" 的区域 R, N_R 是均值为 $\lambda V(R)$ 的泊松分布, 此处 $V(R)$ 是 R 的 "体积", 参数 λ 是固定的并在某种意义下度量分布的强度, 它与 R 的大小和形状无关. 空间的泊松过程是在考虑宇宙空间中星体和星系的分布. 动物和植物的位置分布以及显微镜下玻璃片上细菌的分布等问题中提出来的. 这些思想和概念将在第 16 章中进一步研究.

1.3 一般随机过程的分类

区别随机过程类型的要素是**状态空间**的种类, **指标集** T 以及随机变量 X_t 之间的相关关系.

状态空间 S

状态空间是每个 X_t 所有可能值所在的空间. 若 $S=(0,1,2,\cdots)$, 我们称过程为取自然数值的过程或离散状态空间过程. 若 S 为实直线 $(-\infty,+\infty)$, 我们称 X_t 为实值随机过程. 若 S 是 k 维欧几里得空间, 则称 X_t 为 k 维向量过程.

和单个随机变量的情况类似, 状态空间的选择不一定由所描述的物理现象唯一确定, 但在实际研究中, 人们通常选择其中最合适的一个.

指标集 T

若 $T=(0,1,\cdots)$, 则称 $\{X_t\}$ 为离散时间随机过程, 当 T 离散时, 我们常用 X_n 26 代替 X_t. 若 $T=[0,\infty)$, 则 $\{X_t\}$ 称为连续时间过程.

我们已举例说明指标集 T 可以不是一维的 (如空间泊松过程). 另一个例子是海洋波浪. 我们把纬度和经度坐标作为 t 值, 而 X_t 是位置 t 处的波浪高度.

随机过程的古典类型

我们现在根据 X_t 之间相关关系的不同来划分各种经典类型的随机过程. 如无特别说明, 在例子中均取 $T=[0,\infty)$. 为简单起见, 我们假定随机变量 X_t 取实值.

(a) 平稳独立增量过程

若对任意 n 个实值 $t_1,t_2,\cdots,t_n,t_1<t_2<\cdots<t_n$, 随机变量组 $X_{t_2}-X_{t_1},X_{t_3}-X_{t_2},\cdots,X_{t_n}-X_{t_{n-1}}$ 是独立的, 则称 $\{X_t\}$ 是 **独立增量过程**. 如果指标集 T 包含有最小下标 t_0, 这时我们还假定 $X_{t_0},X_{t_1}-X_{t_0},X_{t_2}-X_{t_1},\cdots,X_{t_n}-X_{t_{n-1}}$ 是独立的. 如果指标集是离散的, 即 $T=(0,1,\cdots)$, 那么有关独立增量过程问题在某种意义下可归结为独立随机变量序列问题. 事实上, 令 $Z_0=X_0,Z_i=X_i-X_{i-1},(i=1,2,\cdots)$, 则 $Z_0,Z_1,\cdots,Z_n,\cdots$ 是独立随机变量序列. 如果我们知道了 $Z_0,Z_1,\cdots,Z_n,\cdots$ 各自的分布函数, 就能确定 (很显然)$\{X_i\}$ 的任何有限维联合分布, 因为

$$X_i=Z_0+Z_1+\cdots+Z_i,\quad i=0,1,2,\cdots.$$

如果增量 $X(t_1+h)-X(t_1)$ 的分布仅仅依赖于区间长度 h 而与 t_1 无关, 则此过程称为**平稳增量**过程. 对于平稳增量过程, 不论怎样的 t_1,t_2 以及 $h,X(t_1+h)-X(t_1)$ 的分布与 $X(t_2+h)-X(t_2)$ 是相同的.

如果过程 $\{X_t,t\in T\}$ 具有平稳独立增量, 且具有有限均值, 这里 $T=[0,\infty)$ 或 27 $T=(0,1,2,\cdots)$, 则可用初等方法证明 $E[X_t]=m_0+m_1t$, 其中 $m_0=E[X_0],m_1=E[X_1]-m_0$. 类似的结论对于方差也成立：

$$\sigma_{X_t}^2=\sigma_0^2+\sigma_1^2t,$$

其中, $\sigma_0^2=E[(X_0-m_0)^2],\sigma_1^2=E[(X_1-m_1)^2]-\sigma_0^2$.

我们将给出均值情形的证明. 设 $f(t)=E[X_t]-E[X_0]$. 则对任何 t 和 s

$$\begin{aligned} f(t+s) &= E[X_{t+s}-X_0] \\ &= E[X_{t+s}-X_s+X_s-X_0] \\ &= E[X_{t+s}-X_s]+E[X_s-X_0] \\ &= E[X_t-X_0]+E[X_s-X_0] \end{aligned}$$

(利用平稳增量的性质)

$$= f(t)+f(s).$$

在适当的条件下, 函数方程 $f(t+s)=f(t)+f(s)$ 只有解 $f(t)=f(1)\cdot t$. 我们将在假定 $f(t)$ 可微的条件下 (显然此条件可进一步削弱) 给出这个结果的证明. 对函数方程两边关于 t 和 s 分别独立地求微分, 得到

$$f'(t+s)=f'(t)=f'(s).$$

因此, 当 $s=1$ 时, 我们有 $f'(t)=f'(1)=c$. 解此初等微分方程得 $f(t)=ct+d$. 但由 $f(0)=2f(0)$ 知 $f(0)=0$, 必有 $d=0$. 由此得 $c=f(1)$. 表达式 $f(t)=f(1)t$ 实际上就是

$$E[X_t]-m_0=(E[X_1]-m_0)\cdot t$$

或

$$E[X_t]=m_0+m_1t,$$

这就是我们要证明的.

布朗运动和泊松过程都是平稳独立增量过程.

(b) 鞅

设 $\{X_t\}$ 是离散指标或连续指标的实值随机过程. 若对于任意的 $t, E[|X_t|]<\infty$, 且对任意 $t_1<t_2<\cdots<t_{n+1}$ 以及任意实数 $a_1,a_2,\cdots,a_n, E\{X_{t_{n+1}}|X_{t_1}=a_1,X_{t_2}=a_2,\cdots,X_{t_n}=a_n\}=a_n$, 则称 $\{X_t\}$ 为鞅. 当 X_t 表示时刻 t 赌徒可拥有的赌本, 鞅可作为描述公平赌博的模型. 鞅的性质表示赌徒在已知时刻 t_n 拥有赌本 a_n 的情况下, 时刻 t_{n+1} 拥有赌本的平均值仍然为 a_n, 而与他在时刻 t_n 以前拥有的赌本无关. 读者容易验证, 当 $Z_i(i=1,2,\cdots)$ 是相互独立且均值为 0 的随机变量时, $X_n=Z_1+Z_2+\cdots+Z_n, n=1,2,\cdots$, 是一个离散时间鞅. 类似地, 若 $X_t, 0\leqslant t<\infty$, 具有独立增量且均值为 0, 则 $\{X_t\}$ 是连续时间鞅 (见初等问题 6).

鞅是第 6 章研究的主要内容.

28

(c) 马尔可夫过程

粗略地说, **马尔可夫过程**是这样一类过程, 如果已知 X_t 的值, 则对任意的 $s, u, s > t > u, X_s$ 的值与 X_u 的值不相关, 即当准确地知道现在情况时, 该过程任何将来的某种性质的概率并不受过去性质所影响. 然而, 当这个过程现在情况还不清楚时, 将来某性质的概率一般要受该系统过去情况所影响. 用数学式子来表达, 所谓马尔可夫过程 $\{X_t\}$, 是指对任意 $t_1 < t_2 < \cdots < t_n < t$,

$$\begin{aligned}\Pr\{a < X_t \leqslant b | X_{t_1} = x_1, X_{t_2} = x_2, \cdots, X_{t_n} = x_n\} \\ = \Pr\{a < X_t \leqslant b | X_{t_n} = x_n\}.\end{aligned} \tag{3.1}$$

设 A 是实直线上一个区间, 函数

$$P(x, s; t, A) = \Pr\{X_t \in A | X_s = x\}, \quad t > s, \tag{3.2}$$

称为**转移概率函数**. 它是研究马尔可夫过程物造的基础. 我们可将条件 (3.1) 改为

$$\Pr\{a < X_t \leqslant b | X_{t_1} = x_1, X_{t_2} = x_2, \cdots, X_{t_n} = x_n\} = P(x_n, t_n; t, A), \tag{3.3}$$

其中 $A = \{\xi | a < \xi \leqslant b\}$. 可以证明, $(X_{t_1}, X_{t_2}, \cdots, X_{t_n})$ 的联合概率分布可由转移函数 (3.2) 和 X_{t_1} 的初始分布函数确定, 稍后 (第 2 章) 在介绍离散时间、离散状态空间的马尔可夫过程时, 将详细说明上述概念.

[29] 状态空间是有限或可数的马尔可夫过程又称为 **马尔可夫链**. 所有样本函数 (或实现) 是连续函数的马尔可夫过程 $\{X_t, t \in [0, \infty)\}$ 称为**扩散过程**. 泊松过程是连续时间的马尔可夫链, 布朗运动是扩散过程.

(d) 平稳过程

一个随机过程 $\{X_t, t \in T\}$(其中 T 可以是 $(-\infty, +\infty)$、$[0, \infty)$、所有整数集合或所有正整数集合) 称为**严格平稳的**, 如果对任意 $h > 0$ 和任意 $t_1, t_2, \cdots, t_n \in T$, 随机变量组

$$(X_{t_1+h}, X_{t_2+h}, \cdots, X_{t_n+h}) \text{ 和 } (X_{t_1}, X_{t_2}, \cdots, X_{t_n})$$

的联合分布函数相同. 这个条件指出该过程依概率意义是均衡的, 我们在哪个具体时刻观察此过程是无关紧要的. 特别地, 对任意时刻 t, X_t 的分布都相同.

称随机过程 $\{X_t, t \in T\}$ 为**宽平稳的**或**协方差平稳**的, 如果每个 X_t 具有有限二阶矩, 并且对任意的 $t, \mathrm{Cov}(X_t, X_{t+h}) = E(X_t X_{t+h}) - E(X_t)E(X_{t+h})$ 仅仅取决于 h. 具有有限二阶矩的严格平稳过程是协方差平稳过程. 同时也存在非严格平稳的协方差平稳过程.

平稳随机过程作为描述许多现象的数学模型是适当的, 例如在通讯理论、天文学、生物学, 甚至在经济学中, 都需要平稳过程. 这方面的详细讨论将在第 9 章进行.

称马尔可夫过程具有平稳转移概率, 如果 $P(x,s;t,A)$(见 (3.2)) 仅是 $t-s$ 的函数. 注意到 $P(x,s;t,A)$ 是已知现在情况的条件概率, 因此没有理由认为具有平稳转移概率的马尔可夫过程是平稳过程. 事实也确实如此.

泊松过程和布朗运动都不是平稳的. 事实上, 不存在非常数的平稳独立增量过程是平稳过程. 然而, 如果 $\{X_t, t\in[0,\infty)\}$ 是布朗运动或泊松过程, 则 $Z_t = X_{t+h} - X_t$ 对任意固定 h 是平稳随机过程.

(e) 更新过程

更新过程是独立同分布正随机变量序列 $\{T_k\}_{k=1,2,\cdots}$, 这里 T_k 表示某“元件”的寿命. 第一个元件在时刻零开始使用, 在时刻 T_1 坏了以后马上换上第二个元件使用, 在时刻 T_1+T_2 损坏后又换第三个元件, 如此继续下去, 因此称其为“更新过程”. 第 n 次更新时间是 $S_n = T_1+T_2+\cdots+T_n$.

30

更新计数过程 $\{N_t\}$ 表示在时间区间 $[0,t]$ 更新的数目 (见图 1-3), 即

$$N_t = n, \quad \text{当 } S_n \leqslant t < S_{n+1}, \quad n = 0,1,2,\cdots.$$

通常我们并不特意区分更新过程及与其关联的更新计数过程, 两者不会发生混淆.

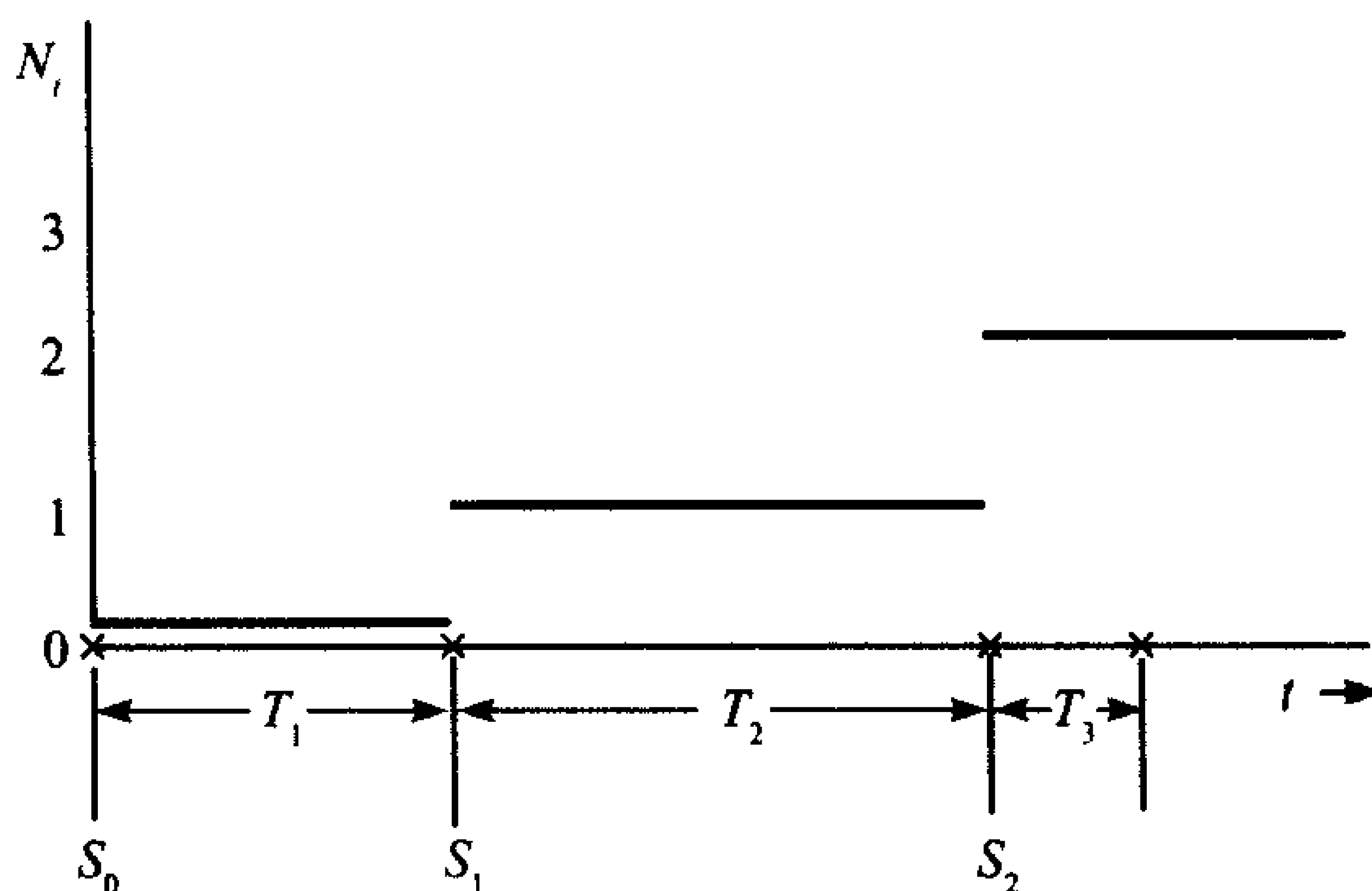


图 1-3

更新过程直接出现在许多应用领域, 诸如管理科学、经济学和生物学等. 另外, 十分重要的是, 其他类型的随机过程初看起来完全与更新过程无关, 却常有更新过程嵌入其中, 第 5 章将致力于研究更新过程.

带有参数 λ 的泊松过程是更新计数过程, 其元件寿命具有参数为 λ 的指数分布.

(f) 点过程

设 S 是 n 维空间的一个集合, $\mathscr{A}$ 是 S 的一个子集族. 点过程是指标集为集合

$A \in \mathscr{A}$ 并以非负整数集 $\{0,1,\cdots,\infty\}$ 为状态空间的随机过程. 我们可以想象"点"以某种随机方式散布在 S 上, $N(A)$ 表示落在集 A 里的点数. 由于 $N(A)$ 是计数函数, 因此对每个实现应附加某些要求. 例如, 如果 A_1 和 A_2 是 $\mathscr{A}$ 中不相交的集合, $A_1 \bigcup A_2$ 也属于 $\mathscr{A}$, 则我们要求

$$N(A_1 \bigcup A_2) = N(A_1) + N(A_2),$$

如果空集 $\varnothing$ 也属于 $\mathscr{A}$, 则 $N(\varnothing) = 0$.

假设 S 是实直线 (平面或三维空间) 的一个集合, 对于每个子集 $A \subset S$, 设 $V(A)$ 表示 A 的长度 (对应地, 面积或体积), 称 $\{N(A), A \subset S\}$ 是强度为 $\lambda > 0$ 的齐次泊松点过程, 如果满足:

31

(i) 对每个 $A \subset S, N(A)$ 为具有参数 $\lambda V(A)$ 的泊松分布;

(ii) 对 S 的 n 个不相交子集 $\{A_1, \cdots, A_n\}$, 随机变量 $N(A_1), \cdots, N(A_n)$ 是独立的.

泊松点过程产生于对宇宙空间星体和星系的分布、植物和动物的平面分布以及细菌分布等问题的研究. 这些思想和概念将在第 2 卷第 16 章进一步研究.

每个泊松过程 $\{X_t, t \in [0,\infty)\}$ 确定 $S = [0,\infty)$ 上的一个泊松点过程. 事实上, 对于任意区间 $A = (s,t], s < t$, 可令 $N(A) = X_t - X_s$.

1.4 随机过程的确定

一个随机过程 $\{X_t\}$ 的特征主要在于随机变量 $X_t (t \in T)$ 之间的相互关系. 这些关系由该过程所有有限个随机变量的联合分布函数所确定. 在本书中, 当过程的指标集. 状态空间以及有限维联合分布函数族被指定, 则认为该过程已被确定. 然而, 在处理连续参数随机过程时都出现了某种困难. 我们用下例来说明.

设 U 是在 $[0,1]$ 上均匀分布的随机变量, 并定义 X_t 和 Y_t 如下:

$$X_t = \begin{cases} 1, & \text{当 } U = t, \\ 0, & \text{其他}, \end{cases}$$
$$Y_t \equiv 0, \qquad (t > 0).$$

通过简单计算可知, $\{X_t\}$ 和 $\{Y_t\}$ 具有相同的有限维分布族. 然而, 显然有

$$\Pr\left\{X_t \leqslant \frac{1}{2}, \text{ 对于所有 } 0 \leqslant t \leqslant 1\right\} = 0$$

和

$$\Pr\left\{Y_t \leqslant \frac{1}{2}, \text{ 对于所有 } 0 \leqslant t \leqslant 1\right\} = 1.$$

这确实使我们感到为难. 为了指出产生此困难的原因, 可考虑下面问题.

假设 $\{X_t, 0 \leqslant t < \infty\}$ 是连续参数随机过程, 我们欲求概率 $\Pr\{X_t \geqslant 0, 0 \leqslant t \leqslant 1\}$.

32

考虑递减事件序列

$$A_n = \{X_{t_i} \geqslant 0, t_i = i/2^n, i = 0, 1, \cdots, 2^n\}, n = 1, 2, \cdots.$$

每个 A_n 的概率可以利用 $X_{t_i}, i = 0, 1, \cdots, 2^n$, 的联合分布函数来计算, 然后把极限值 $\lim_{n\to\infty} \Pr\{A_n\}$ 作为 $\Pr\{X_t \geqslant 0, 0 \leqslant t \leqslant 1\}$. 表面上看来这似乎是合理的. 但是, 这种表面上的合理性却未必无矛盾. 我们考虑事件列

$$A'_n = \{X_{t_i} \geqslant 0, t_i = i/3^n, i = 0, 1, \cdots, 3^n\},$$

取极限值 $\lim_{n\to\infty} \Pr\{A'_n\}$ 作为 $\Pr\{X_t \geqslant 0, 0 \leqslant t \leqslant 1\}$ 似乎同样合情合理, 可是等式 $\lim_{n\to\infty} \Pr\{A_n\} = \lim_{n\to\infty} \Pr\{A'_n\}$ 绝不是显然的. 事实上, 若无随机过程样本函数 "光滑性" 的假设, 此二极限值并不一定相等. 要使这两个极限相等, 有若干不同的充分条件, 其中之一是, 对任意 $\varepsilon > 0$ 和任意 $t, \lim_{\tau\to t} \Pr\{|X_t - X_\tau| \geqslant \varepsilon\} = 0$. 在这个条件下, 我们把 $\Pr\{X_t \geqslant 0, 0 \leqslant t \leqslant 1\}$ 定义为这两个极限的共同值就不会引起矛盾. 实际上, 若 $t_1, t_2, \cdots$ 是区间 $[0, 1]$ 上的稠密集, 则 $\Pr\{X_t \geqslant 0, 0 \leqslant t \leqslant 1\} = \lim_{n\to\infty} \Pr\{X_{t_i} \geqslant 0, i = 1, 2, \cdots, n\}$.

上述问题的症结在于, 根据全概率公理, 关于随机变量序列事件的概率虽然可以用含有该序列的有限子集 (因而是可数子集) 的事件的概率来计算, 但是, 现在我们所遇到的事件 $\{X_t \geqslant 0, 0 \leqslant t \leqslant 1\}$ 都包含了不可数个随机变量. 关于这方面问题的详细讨论是十分复杂的, 已超出本书的范围, 有兴趣的读者可参考 Doob 书的第 2 章.[1]有助于理解这方面问题的一些基本事实将在本书第 14 章予以讨论.

初等问题

1. 设 X 是可能值为 0, 1, 2, $\cdots$ 的非负离散随机变量, 证明

$$E[X] = \sum_{n=0}^{\infty} \Pr\{X > n\} = \sum_{k=1}^{\infty} \Pr\{X \geqslant k\}.$$

提示: $E[X] = \sum_{n=1}^{\infty} n\Pr\{X = n\} = \sum_{n=1}^{\infty}\sum_{k=1}^{n} \Pr\{X = n\}$.

33

2. 假设一个罐子里有几张标有号码 $1, 2, \cdots, n$ 的纸片, 某人抽出其中一张后放回, 再抽一张随后又放回, 一直到他抽到先前已经抽过的纸片时停止. 令 X 表示达到目的时抽取的次数. 试求 X 的概率分布.

1. J.L. Doob, *Stochastic Processes*, Wiley, New York, 1953.

提示：先计算 $\Pr\{X > k\}$ 最容易.

答案：$p(k) = (k-1)!\binom{n}{k-1}\frac{k-1}{n^k}$, 对于 $k = 2, 3, \cdots, n+1$.

3. 试证问题 2 中随机变量 X 的期望是

$$E(X) = 2 + \left(1 - \frac{1}{n}\right) + \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) + \cdots + \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{n-1}{n}\right).$$

提示：利用初等问题 1.

4. 某工厂在一周内出现事故的次数是均值为 μ、方差为 σ^2 的随机变量, 在不同事故中损坏的个体数目彼此独立地服从均值为 ν、方差为 τ^2 的分布. 试确定在一周内损坏的个体数目的均值和方差.

答案：均值为 $\mu\nu$, 方差为 $\nu^2\sigma^2 + \mu\tau^2$.

5. 若我们按参数为 λ 的泊松随机变量 X 进行观测, 此时, 成功概率为 p 的二项事件重复 X 次, 并且观测到 σ 次成功, 试问 σ 的分布是什么?

提示：利用随机变量随机和的母函数.

答案：泊松分布, 其参数为 λp.

6. 证明：均值为 0 的独立随机变量 X_n 之和 $S_n = X_1 + \cdots + X_n$ 构成一个鞅. 这里假设 $E[|X_k|] < \infty$, 对于 $k = 1, 2, \cdots$.

7. 证明带有独立增量的随机过程 $\{X(t), t = 0, 1, 2, \cdots\}$ 是马尔可夫过程. (注：此结论对于随机过程 $\{X(t), -\infty < t < \infty\}$ 不成立.)

8. 设有 n 对夫妇, 其中第 i 对夫妇生男孩的概率为 p_i, 并且孩子数的期望值为 c_i. 假设所有各对的 p_i 都与时间无关, 任一对相继出生的孩子的性别是独立的随机变量, 并且假设没有多胞胎出生. 定义性别比为

34

$$S = \frac{n\text{ 对夫妇出生男孩的期望数}}{n\text{ 对夫妇出生孩子的期望数}}.$$

假定 $c_i = c, i = 1, 2, \cdots, n$. 求 S.

答案：$S \equiv S_0 = \left(\sum_{i=1}^{n} p_i\right) \Big/ n$.

(以下问题 9~13 的条件同问题 8.)

9. 若所有双亲决定, 生出一个男孩就停止生育. 试证

$$S = S_1 = \frac{n}{\sum_{i=1}^{n} \frac{1}{p_i}} \leqslant S_0.$$

10. 假设所有双亲决定, 若他们的第一个孩子是男孩, 则继续生育直到生女孩子为止; 若他们的第一个孩子是女孩, 则继续生育直到生男孩为止. 试计算在这种生育计划之下的 S.

答案:

$$S=S_2=\frac{\left\{\sum_{i=1}^{n}1/q_i-\sum_{i=1}^{n}p_i\right\}}{\sum_{i=1}^{n}[p_iq_i]^{-1}-n},$$

其中 $q_i=1-p_i$.

11. 假如所有双亲决定, 若他们的第一个孩子是男孩, 则继续生育直到生女孩为止; 若他们的第一个孩子是女孩就不再生了. 试计算对应的 S.

答案:

$$S=S_3=1-\left[\frac{n}{\sum_{i=1}^{n}1/q_i}\right].$$

12. 证明: 或者 $S_2\leqslant S_0$, 或者 $S_2>S_0$, 这取决于 $\{p_1,p_2,\cdots,p_n\}$ 的值, 其中 S_0 是初等问题 8 的性别比, S_2 是初等问题 10 的性别比.

13. (a) 假设 n 对双亲中第 i 对双亲生的孩子得病的概率是 $p_i, i=1,2,\cdots,n$. 若有一个病孩出生则其双亲不敢再生育. 令 s = 一个家庭没有病孩出生时其子女的数目. 假定 p_i 与前面的出生情况无关. 试证明

35

$$R_1=\frac{\text{有病孩子的期望数}}{\text{出生孩子总数的期望值}}=\frac{\sum_{i=1}^{n}(1-q_i^s)}{\sum_{i=1}^{n}(1-q_i^s)/p_i}.$$

(b) 假定两个病孩 (而不是一个) 的出生将使其双亲不敢再生育. 证明在这种限制之下

$$R_2=\frac{\sum_{i=1}^{n}\{2(1-q_i^s)-sp_iq_i^{s-1}\}}{\sum_{i=1}^{n}\{[2(1-q_i^s)/p_i]-sq_i^{s-1}\}}.$$

问 题

下述积分在某些问题中可能有用, 现列于此, 以供参考.

定义伽玛函数为

$$\Gamma(x)=\int_0^{\infty}\xi^{x-1}\mathrm{e}^{-\xi}\mathrm{d}\xi,\quad x>0.$$

对于较大的 $x, \Gamma(x)\sim\sqrt{2\pi}\mathrm{e}^{-x}x^{x+1/2}$(斯特林公式). 当 $x=n$(正整数) 时, $\Gamma(n)=(n-1)!=(n-1)(n-2)\cdots 2\times 1$.

Beta 积分定义为

$$\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}=\int_0^1 x^{p-1}(1-x)^{q-1}\mathrm{d}x,$$

其中, $p>0,q>0$.

1. 设 a,b,c 是 $(0,1)$ 上均匀分布的独立随机变量. 试求 ax^2+bx+c 有实根的概率.

答案: $(5+3\ln 4)/36$.

2. 对于每个固定的 $\lambda>0$, 设 X 具有参数为 λ 的泊松分布, 假设 λ 本身是服从 Γ 分布的随机变量 (即具有密度

36

$$f(\lambda)=\begin{cases}\dfrac{1}{\Gamma(n)}\lambda^{n-1}\mathrm{e}^{-\lambda}, & \lambda\geqslant 0,\\ 0, & \lambda<0,\end{cases}$$

其中 n 是固定正常数). 试证明

$$\Pr\{X=k\}=\frac{\Gamma(k+n)}{\Gamma(n)\Gamma(k+1)}\left(\frac{1}{2}\right)^{k+n},\quad k=0,1,\cdots.$$

当 n 是一个整数时, 这是 $p=\dfrac{1}{2}$ 的负二项分布.

3. 对每个固定 p, 令 X 具有参数为 p 和 N 的二项分布. 假设 N 本身具有参数为 q 和 $M(M\geqslant N)$ 的二项分布

(a) 解析地证明 X 具有参数为 pq 和 M 的二项分布.

(b) 对此结果给予概率上的说明.

4. 对每个固定 p, 令 X 具有参数为 p 和 N 的二项分布. 假设 N 本身具有参数为 r 和 s 的 Beta 分布. 试求复合结果 X 的分布. 何时此分布为 $x=0,1,\cdots,N$ 上的均匀分布?

答案: $\Pr\{X=k\}=\dbinom{N}{k}\dfrac{\Gamma(r+s)\Gamma(k+r)\Gamma(N-k+s)}{\Gamma(r)\Gamma(s)\Gamma(N+r+s)}$;

$$\Pr\{X=k\}=\frac{1}{N+1},\quad 当\ r=s=1\ 时.$$

5. (a) 设 X 服从参数为 λ 的泊松分布, 参数 λ 本身是随机变量, 服从均值为 $1/c$ 的指数分布. 求 X 的分布.

(b) 如果 λ 服从阶数为 α, 比例参数为 c 的 Γ 分布, 即 λ 的密度是

$$c^{\alpha+1}\frac{\lambda^{\alpha}}{\Gamma(\alpha+1)}\mathrm{e}^{-\lambda c},$$

对 $\lambda>0$; 其余为 0. 此时 X 的分布如何?

答案: (a) $\Pr\{X=k\}=\dfrac{c}{(c+1)^{k+1}}$;

(b) $\Pr\{X=k\}=\dfrac{\Gamma(k+\alpha+1)}{k!\Gamma(\alpha+1)}\left(\dfrac{1}{1+c}\right)^{k+\alpha+1}c^{\alpha+1}$.

6. 假设我们有 N 个分别标有 $1,2,\cdots,N$ 的小片, 我们无返回地抽取容量为 $2n+1$ 的随机样本, 设 Y 是随机样本的中位数. 证明 Y 的分布律是

$$\Pr\{Y=k\}=\frac{\dbinom{k-1}{n}\dbinom{N-k}{n}}{\dbinom{N}{2n+1}},\quad 对于\ k=n+1,n+2,\cdots,N-n.$$

验证

$$E[Y]=\frac{N+1}{2}\text{ 和 }\operatorname{Var}(Y)=\frac{(N-2n-1)(N+1)}{8n+12}.$$

37

7. 假设我们有 N 个分别标有 $1,2,\cdots,N$ 的小片. 无返回地抽取容量为 n 的随机样本, 设 X 是随机样本的最大数. 证明 X 的分布律是

$$\Pr\{X=k\}=\frac{\binom{k-1}{n-1}}{\binom{N}{n}},\quad \text{对于 } k=n,n+1,\cdots,N.$$

验证

$$E[X]=\frac{n}{n+1}(N+1),\quad \operatorname{Var}(X)=\frac{n(N-n)(N+1)}{(n+1)^2(n+2)}.$$

8. 设 X_1 和 X_2 是在区间 $\left[\theta-\frac{1}{2},\theta+\frac{1}{2}\right]$ 上均匀分布的独立随机变量. 证明 X_1-X_2 的分布与 θ 无关, 并求此分布的密度函数.

答案:

$$f_{X_1-X_2}(y)=\begin{cases}1+y, & -1\leqslant y\leqslant 0,\\ 1-y, & 0<y<1,\\ 0, & |y|\geqslant 1.\end{cases}$$

9. 设 X 是非负随机变量, 分布函数为 $F(x)=\Pr\{X\leqslant x\}$. 证明

$$E[X]=\int_0^\infty[1-F(x)]\mathrm{d}x.$$

提示: $E[X]=\int_0^\infty x\mathrm{d}F(x)=\int_0^\infty\left(\int_0^x\mathrm{d}y\right)\mathrm{d}F(x)$.

10. 设 X 是非负随机变量并令

$$\begin{aligned}X_c&=\min\{X,c\}\\&=\begin{cases}X, & \text{若 } X\leqslant c,\\ c, & \text{若 } X>c,\end{cases}\end{aligned}$$

其中 c 是已知常数, 试利用分布函数 $F(x)=\Pr\{X\leqslant x\}$ 表示期望 $E[X_c]$.

答案: $E[X_c]=\int_0^c[1-F(x)]\mathrm{d}x$.

11. 设 X 和 Y 是具有可能值 $0,1,2,\cdots$ 的离散型随机变量, 对于 $|s|<1,|t|<1$, 定义联合母函数

$$\phi_{X,Y}(s,t)=\sum_{i,j=0}^\infty s^it^j\Pr\{X=i,Y=j\}$$

38

和边沿母函数

$$\phi_X(s) = \sum_{i=0}^{\infty} s^i \Pr\{X = i\},$$

$$\phi_Y(t) = \sum_{j=0}^{\infty} t^j \Pr\{Y = j\}.$$

(a) 试证明, X 和 Y 是独立的当且仅当

$\phi_{X,Y}(s,t) = \phi_X(s) \cdot \phi_Y(t)$, 对于所有 s, t.

(b) 试给出随机变量 X 和 Y 的具体例子, 这里 X 和 Y 不独立, 但满足

$$\phi_{X,Y}(t,t) = \phi_X(t) \cdot \phi_Y(t) \quad \text{对于所有的 } t.$$

(注意到 $\phi_{X,Y}(t,t)$ 是 $X+Y$ 的母函数, 因此, 此例说明虽然由独立性可推得随机变量和的母函数是其边缘母函数的乘积, 但反之不成立.)

12. 现有作为某试验结果的 $r+1$ 个事件 $A_0, A_1, \cdots, A_r$. 令 p_i 是 $A_i (i = 0, 1, , 2, \cdots, r)$ 出现的概率. 若我们独立地进行试验直至事件 A_0 出现 k 次为止. 设 X_i 是事件 A_i 出现的次数. 试证明

$$\Pr\left\{X_1 = x_1, \cdots, X_r = x_r; A_0 \text{ 在 } \left(k + \sum_{i=1}^{r} x_i\right) \text{ 次试验中出现 } k \text{ 次}\right\}$$
$$= \frac{\Gamma\left(k + \sum_{i=1}^{r} x_i\right)}{\Gamma(k) \prod_{i=1}^{r} x_i!} p_0^k \prod_{i=1}^{r} p_i^{x_i}. \tag{I}$$

13. 证明带有参数 $(k; p_0, p_1, \cdots, p_r)$ 的**负多项分布**(I) 的概率母函数是

$$\varphi(t_1, \cdots, t_r) = p_0^k \left(1 - \sum_{i=1}^{r} t_i p_i\right)^{-k}.$$

14. 考虑随机向量 $\{X_0, X_1, \cdots, X_r\}$, 它服从参数为 $(n; p_0, p_1, \cdots, p_r)$ 的多项分布; 并设 n 本身是一个随机变量. 服从参数为 (k, p) 的负二项分布. 试求 $X_0, X_1, \cdots, X_r$ 的联合分布.

15. 设一批签分成 $0, 1, \cdots, r$ 等级, 每一等级分别有 $m, n_1, \cdots, n_r$ 条. 我们一条接一条无返回地抽取, 直到第 0 等级签抽满 k 条时停止. 试证第 $1, \cdots, r$ 等级的签被抽到的频数 $X_1, \cdots, X_r$ 的联合分布为

39

$$\Pr\{X_1 = x_1, \cdots, X_r = x_r\}$$
$$= \left\{\binom{m}{k-1} \prod_{i=1}^{r} \binom{n_i}{x_i} \Big/ \binom{m+n}{k+y-1}\right\} \cdot \frac{m-(k-1)}{m+n-(k+y-1)},$$

其中, $y = \sum_{i=1}^{r} x_i, n = \sum_{i=1}^{r} n_i$.

16. 继续问题 15. 如果 $m \to \infty$ 和 $n \to \infty$ 使得 $m/(m+n) \to p_0$ 和 $n_i/(m+n) \to p_i, i=1,2,\cdots,r$, 证明问题 15 的分布逼近负多项分布.

17. 随机变量 X_n 取值 $\frac{k}{n}, k=1,2,\cdots,n$, 取各值的概率均为 $\frac{1}{n}$. 试求它的特征函数及 $n \to \infty$ 时的极限, 然后求极限特征函数对应的随机变量.

答案: (a) $\varphi_n(t) = (1-\mathrm{e}^{\mathrm{i}t})\frac{1}{n} \cdot \frac{1}{\exp(-\mathrm{i}n^{-1}t)-1}$;

(b) 在 $(0,1)$ 的均匀分布.

18. 利用对于适当的泊松随机变量的中心极限定理证明

$$\lim_{n\to\infty} \mathrm{e}^{-n}\sum_{k=0}^{n}\frac{n^k}{k!} = \frac{1}{2}.$$

*19. 设随机变量 X 和 Y 具有如下性质: X 是正的, 即 $\Pr\{X>0\}=1$, 且有连续密度函数 $f(x), Y|X$ 服从 $\{0,X\}$ 上的均匀分布. 试证明, 如果 Y 和 $X-Y$ 是独立的, 则

$$f(x) = a^2 x\mathrm{e}^{-ax}, \quad x>0, a>0.$$

*20. 设 U 是阶数为 p 的 Γ 分布, V 具有参数为 q 和 $p-q(0<q<p)$ 的 Beta 分布. 假定 U 和 V 独立, 试证明 UV 是阶数为 q 的 Γ 分布.

提示:

$$\Pr\{UV \leqslant x\} = \int_0^1 \left(\int_0^{x/\xi} \mathrm{e}^{-\lambda}\lambda^{p-1}\mathrm{d}\lambda\right) \frac{\xi^{q-1}(1-\xi)^{p-q-1}\mathrm{d}\xi}{\Gamma(q)\Gamma(p-q)}.$$

两边取拉普拉斯变换, 改变积分顺序, 然后利用幂级数展开式

$$(1-y)^{-\alpha-1} = \sum_{k=0}^{\infty}\binom{\alpha+k}{k}y^k.$$

40

*21. 设 X 和 Y 是有连续密度函数 $f(x)$ 的独立同分布的正随机变量, 且 $f(x)>0$. 进一步假设 $U=X-Y$ 和 $V=\min(X,Y)$ 是独立的随机变量. 证明对某个 $\lambda>0$ 成立

$$f(x) = \begin{cases} \lambda\mathrm{e}^{-\lambda x}, & x \geqslant 0, \\ 0, & \text{其他}. \end{cases}$$

提示: 首先证明 U 和 V 的联合密度函数是

$$f_{U,V}(u,v) = f(v)\cdot f(v+|u|).$$

然后使其等于 U 和 V 的边缘密度的乘积.

22. 设 X 和 Y 是两个相互独立、取非负整数值的随机变量, 对于所有非负整数 x 和 y,

$$\Pr\{X=x|X+Y=x+y\} = \frac{\dbinom{m}{x}\dbinom{n}{y}}{\dbinom{m+n}{x+y}},$$

其中 m 和 n 是已知的正整数. 假设 $\Pr\{X=0\}$ 和 $\Pr\{Y=0\}$ 是严格正的. 证明 X 和 Y 具有相同参数 p 的二项分布, 另一参数分别为 m 和 n.

23. (a) 设 X 和 Y 是独立的随机变量, 满足

$$\Pr\{X=i\}=f(i), \quad \Pr\{Y=i\}=g(i),$$

其中

$$f(i)>0, \quad g(i)>0, \quad i=0,1,2,\cdots,$$

且

$$\sum_{i=0}^{\infty} f(i)=\sum_{i=0}^{\infty} g(i)=1.$$

假设

$$\Pr\{X=k|X+Y=l\}=\begin{cases}\binom{l}{k}p^k(1-p)^{l-k}, & 0\leqslant k\leqslant l,\\ 0, & k>l.\end{cases}$$

证明

$$f(i)=\mathrm{e}^{-\theta\alpha}\frac{(\theta\alpha)^i}{i!}, \quad g(i)=\mathrm{e}^{-\theta}\frac{\theta^i}{i!}, \quad i=0,1,2,\cdots,$$

其中, $\alpha=p/(1-p), \theta>0$ 是任意的.

(b) 证明 p 由条件

$$G\left(\frac{1}{1-p}\right)=\frac{1}{f(0)}$$

确定.

提示: 令 $F(s)=\sum f(i)s^i, G(s)=\sum g(i)s^i$. 首先证明关系式

41

$$F(u)F(v)=F(vp+(1-p)u)G(vp+(1-p)u).$$

24. 设 X 是非负整数值随机变量, 其概率母函数为 $f(s)=\sum_{n=0}^{\infty}a_n s^n$. 通过观测 X 次, 导出成功概率为 p 的 X 次二项试验. 设 Y 表示复合结果成功次数.

(a) 试确定 Y 的概率母函数.

(b) 试确定在已知 $Y=X$ 的条件下 X 的概率母函数.

答案: (a) $f(1-p+ps)$; (b) $f(ps)/f(p)$.

25. (问题 24 的继续) 设对每个 $p(0<p<1)$, (a) 和 (b) 的概率母函数相同, 试证明 X 的分布是泊松的, 即对某个 $\lambda>0$ 有 $f(s)=\mathrm{e}^{\lambda(s-1)}$.

26. 关于股票价格变化的统计分布, 或更一般地, 关于股票价格序列的随机模型, 至少存在四个不同学派. 就其追随者人数而言, 非常流行的一种方法是所谓的“市场分析”, 使用了诸如短期趋势, 维持和忍受水平, 市场反应, 等等. 与这种专业化的观点不同的另外两个学派认为价格序列是一个随机游动, 价格的变化与以前价格是统计独立的. 但这两个学派在选择适当的概率分布方面意见有分歧. 一些人认为价格的变化服从正态分布, 而另一些人认

为此分布应具有“重尾概率”，甚至具有无限的方差. 第四个学派 (与前两个有交叉) 赞成把随机游动作为第一级逼近, 但指出第二级的影响.

这道练习题欲说明上述中间两个学派观点之间的协调性. 人们已经注意到, 那些认为价格变化是依正态规律的人是在固定的交易次数中测量价格的改变, 而那些认为价格改变具有重尾概率的人是在固定时间区间内测量价格的改变, 在此期间进行的交易次数是随机的. 设 Z 是价格的变化量. 用来度量“重尾性” (对此尚有争议) 的是超出量:

$$\gamma_2 = \left[m_4/(m_2)^2\right] - 3$$

的系数, 其中 m_k 是 Z 关于其均值的第 k 阶矩.

假设在每次交易中价格以相等的概率上升一个单位或下降一个单位, N 表示交易的次数, 记 $Z = X_1 + X_2 + \cdots + X_N$, 其中 X_n 是独立同分布的随机变量, 且等可能地取值 $+1$ 和 -1. 试对 Z 计算 γ_2: (a) 当 N 是一个固定数 a; (b) 当 N 是服从均值为 a 的泊松分布.

27. 设有无限多数目的罐, 我们独立地往罐里抛球, 使得任一球落在第 k 个罐里的概率是 $1/2^k, k = 1, 2, 3, \cdots$. 对每个正整数 N, 令 Z_N 表示在第 N 次抛球之后至少含有一个球的罐的数目. 试证明

42

$$E(Z_N) = \sum_{k=1}^{\infty}\left[1 - \left(1 - \frac{1}{2^k}\right)^N\right],$$

并且存在常数 $C_1 > 0$ 和 $C_2 > 0$, 使得对于所有的 N,

$$C_1 \ln N \leqslant E(Z_N) \leqslant C_2 \ln N$$

提示: 验证并利用以下事实:

$$E(Z_N) \geqslant \sum_{k=1}^{\ln_2 N}\left[1 - \left(1 - \frac{1}{2^k}\right)^N\right] \geqslant C \ln_2 N$$

及

$$1 - \left(1 - \frac{1}{2^k}\right)^N \leqslant N\frac{1}{2^k} \quad 与 \quad N\sum_{\ln_2 N}^{\infty}\frac{1}{2^k} \leqslant C_2.$$

28. 设 L 和 R 是随机选择的区间端点, 且具有任意的联合分布, 但应满足 (这是自然的)$L \leqslant R$. 令 $p(x) = \Pr\{L \leqslant x \leqslant R\}$ 是区间覆盖点 x 的概率, $X = R - L$ 是区间长度, 试证公式 $E[X] = \int_{-\infty}^{\infty} p(x)\mathrm{d}x$.

29. 设 N 个球独立向 n 个罐抛掷, 每个球落入任一罐的概率均为 $\frac{1}{n}$. 令 $Z_{N,n}$ 表示全部抛掷之后空罐的数目, 记 $P_{N,n}(k) = \Pr(Z_{N,n} = k)$, 定义 $\varphi_{N,n}(t) = \sum_{k=0}^{n} P_{N,n}(k)\mathrm{e}^{\mathrm{i}kt}$.

(a) 证明

$$P_{N+1,n}(k) = \left(1 - \frac{k}{n}\right)P_{N,n}(k) + \frac{k+1}{n}P_{N,n}(k+1), \text{ 对于 } k = 0, 1, \cdots, n.$$

(b) 证明

$$P_{N,n}(k) = \left(1-\frac{1}{n}\right)^N P_{N,n-1}(k-1) + \sum_{i=1}^{N}\binom{N}{i}\frac{1}{n^i}\left(1-\frac{1}{n}\right)^{N-i} P_{N-i,n-1}(k).$$

(c) 定义 $G_n(t,z) = \sum_{N=0}^{\infty}\varphi_{N,n}(t)\cdot\frac{n^N}{N!}z^N$. 利用 (b), 证明 $G_n(t,z) = G_{n-1}(t,z)(\mathrm{e}^{\mathrm{i}t}+\mathrm{e}^z-1)$, 从而推出

43

$$G_n(t,z) = (\mathrm{e}^{\mathrm{i}t}+\mathrm{e}^z-1)^n, \quad n = 0,1,2,\cdots.$$

附 记

关于概率论及其应用方面丰富多彩的介绍见 Feller[1]. 但 Feller 的书局限于讨论离散概率.

Gnedenko 的教科书 [2] 也是一本优秀的入门书.

另一本有益的初级教科书是 Parzen 的 [3].

关于随机过程的经典著作是 Doob 的 [4]. 该书对研究随机过程是必不可少的.

另一本关于随机过程构造的杰出的书是 Dynkin [5] 的最近翻译本.

参考书目

[1] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol 1, 2nd ed. Wiley, New York, 1957. (有中译本《概率论及其应用 (第 3 版)》, 胡迪鹤译, 人民邮电出版社出版.)

[2] B. V. Gnedenko, *Theory of Probability*, Chelsea, New York, 1962. (有中译本, 丁寿仁译, 1954 年第 2 版).

[3] E. Parzen, *Modern Probability Theory and Its Applications*. Wiley, New York, 1960.

[4] J. L. Doob, *Stochastic Processes*. Wiley, New York, 1953.

44

[5] E. B. Dynkin, *Theory of Markov Processes*, Academic Press, New York, 1965.

第2章　马尔可夫链

这一章介绍马尔可夫链，其内容可在任何一本随机过程初级教程中找到. 本章首先复习马尔可夫链的精确定义 (读者可参阅 1.3 节中的例 (c)).

读者可尝试构造例子解释 2.4 节提到的**可达到**、**互通**、**非周期**、**常返**、**瞬态**和**不可约**等概念.

2.7 节仅有一页, 初次阅读可略过.

2.1　定　　义

离散时间马尔可夫链 $\{X_n\}$ 是状态空间可数或有限且 $T=(0,1,2,\cdots)$ 的马尔可夫随机过程. 我们可以把 X_n 的值看作第 n 次试验的结果.

为了方便, 通常用非负整数集 $\{0,1,2,\cdots\}$ 标记这个过程的状态空间. 当 $X_n=i$ 时, 习惯上就称 X_n 处于状态 i.

已知 X_n 为状态 i, X_{n+1} 到达状态 j 的概率 (称为单步转移概率), 用 $P_{ij}^{n,n+1}$ 来表示, 即

$$P_{ij}^{n,n+1}=\Pr\{X_{n+1}=j|X_n=i\}. \tag{1.1}$$

上述记号强调, 一般地, 转移概率不仅仅是初始状态和最后状态的函数, 而且也是转换时间的函数. 若单步转移概率与时间 (即 n) 无关, 则我们称马尔可夫过程具有*平稳转移概率*(见 1.3 节末). 由于所遇见的马尔可夫链大部分具有平稳转移概率, 所以我们主要讨论这种情形.

在这种情况下, $P_{ij}^{n,n+1}=P_{ij}$ 是与 n 无关的, P_{ij} 表示在一次试验中状态从 i 转移到 j 的概率. 习惯上, 我们把 P_{ij} 列成如下无限方阵的形式

$$\boldsymbol{P}=\left\|\begin{matrix} P_{00} & P_{01} & P_{02} & P_{03} & \cdots \\ P_{10} & P_{11} & P_{12} & P_{13} & \cdots \\ P_{20} & P_{21} & P_{22} & P_{23} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \\ P_{i0} & P_{i1} & P_{i2} & P_{i3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{matrix}\right\|$$

并称 $\boldsymbol{P}=\|P_{ij}\|$ 为这个过程的马尔可夫矩阵或*转移概率矩阵*.

矩阵 $\boldsymbol{P}$ 的第 $i+1$ 行是在已知 $X_n=i$ 的条件下 X_{n+1} 的概率分布. 若状态空间状态数目是有限的, 则 $\boldsymbol{P}$ 是有限方阵且它的阶等于状态的个数. 显然, P_{ij} 满足条件

$$P_{ij} \geqslant 0, \qquad i,j=0,1,2,\cdots, \tag{1.2}$$

$$\sum_{j=0}^{\infty} P_{ij}=1, \qquad i=0,1,2,\cdots. \tag{1.3}$$

条件 (1.3) 仅仅表示在每次试验中必定会出现某个转移. (为方便起见, 即使经过一次试验之后状态保持不变, 我们仍然说发生了一次转移).

一旦 (1.1) 和 X_0 的值 (或者更一般地说, X_0 的分布) 被指定, 过程就被完全确定. 下面我们来证明这个事实.

设 $\Pr\{X_0=i\}=p_i$. 只要指出如何计算概率

$$\Pr\{X_0=i_0, X_1=i_1, X_2=i_2, \cdots, X_n=i_n\}, \tag{1.4}$$

即可, 依照全概率公理可知, 涉及 $X_{j_1}, X_{j_2}, \cdots, X_{j_k}, j_1<j_2<\cdots<j_k$ 的任何概率都是形如式 (1.4) 的各项之和.

由条件概率定义, 我们得到

$$\begin{aligned}
&\Pr\{X_0=i_0, X_1=i_1, X_2=i_2, \cdots, X_n=i_n\} \\
=&\Pr\{X_n=i_n|X_0=i_0, X_1=i_1, \cdots, X_{n-1}=i_{n-1}\}\cdot \\
&\Pr\{X_0=i_0, X_1=i_1, \cdots, X_{n-1}=i_{n-1}\}.
\end{aligned} \tag{1.5}$$

现在根据马尔可夫过程的定义,

$$\begin{aligned}
&\Pr\{X_n=i_n|X_0=i_0, X_1=i_1, \cdots, X_{n-1}=i_{n-1}\} \\
=&\Pr\{X_n=i_n|X_{n-1}=i_{n-1}\}=P_{i_{n-1},i_n}.
\end{aligned} \tag{1.6}$$

把式 (1.6) 代入式 (1.5) 得

$$\begin{aligned}
&\Pr\{X_0=i_0, X_1=i_1, \cdots, X_n=i_n\} \\
=&P_{i_{n-1},i_n}\cdot\Pr\{X_0=i_0, X_1=i_1, \cdots, X_{n-1}=i_{n-1}\}.
\end{aligned} \tag{1.7}$$

由归纳法, 式 (1.4) 变为

$$\begin{aligned}
&\Pr\{X_0=i_0, X_1=i_1, \cdots, X_n=i_n\} \\
=&P_{i_{n-1},i_n}P_{i_{n-2},i_{n-1}}\cdots P_{i_0,i_1}P_{i_0}.
\end{aligned} \tag{1.8}$$

2.2 马尔可夫链的例子

马尔可夫链的重要性在于有大量物理、生物和经济现象可以用它来描述. 下面我们略举数例.

A. 空间齐次马尔可夫链

设 ξ 表示仅取非负整数值的离散值随机度量, $\Pr\{\xi = i\} = a_i, a_i \geqslant 0$, 并且 $\sum_{i=1}^{\infty} a_i = 1$. 设 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ 表示对 ξ 的一系列独立观察.

下面我们描述与 $\xi_i (i = 1, 2, \cdots)$ 相联系的两个不同的马尔可夫链.

在每一种情况下, 其状态空间都是非负整数集合.

(i) 考虑过程 $X_n, n = 0, 1, 2, \cdots$, 定义 $X_n = \xi_n (X_0 = \xi_0$ 事先指定). 它的马尔可夫矩阵有如下形式

$$\boldsymbol{P} = \left\| \begin{matrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ \vdots & & & & \end{matrix} \right\|.$$

矩阵的各行对应值恒等表明随机变量 X_{n+1} 与 X_n 是独立的.

(ii) 另一类重要的马尔可夫链源于对 ξ_i 的部分和序列 η_n 的考虑, 即

$$\eta_n = \xi_1 + \xi_2 + \cdots + \xi_n, \quad n = 1, 2, \cdots,$$

且定义 $\eta_0 = 0$. 显而易见过程 $X_n = \eta_n$ 是马尔可夫链. 我们容易计算它的转移概率矩阵为

47

$$\begin{aligned} & \Pr\{X_{n+1} = j | X_n = i\} \\ = & \Pr\{\xi_1 + \xi_2 + \cdots + \xi_{n+1} = j | \xi_1 + \xi_2 + \cdots + \xi_n = i\} = \Pr\{\xi_{n+1} = j - i\} \\ = & \begin{cases} a_{j-i}, & \text{对于} j \geqslant i, \\ 0, & \text{对} j < i, \end{cases} \end{aligned}$$

这里我们利用了 ξ_i 的独立性假设.

写成矩阵形式为

$$\boldsymbol{P} = \left\| \begin{matrix} a_0 & a_1 & a_2 & a_3 & a_4 & \cdots \\ 0 & a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & 0 & a_0 & a_1 & a_2 & \cdots \\ \vdots & & & & & \end{matrix} \right\|. \tag{2.1}$$

如果随机变量 ξ 允许取正整数和负整数, 则 η_n 可能取遍全体整数. 取代用非负整数集代表状态空间的惯常方法, 这时用整数全体代表状态空间会更方便些, 因为这样其概率转移矩阵将呈现更加对称的形式. 于是状态空间由数 $\cdots,-2,-1,0,1,2,\cdots$ 组成, 其转移概率矩阵为

$$
\boldsymbol{P}=\left\|\begin{array}{ccccccc}
 & \vdots & \vdots & \vdots & \vdots & \vdots & \\
\cdots & a_{-1} & a_0 & a_1 & a_2 & a_3 & \cdots \\
\cdots & a_{-2} & a_{-1} & a_0 & a_1 & a_2 & \cdots \\
\cdots & a_{-3} & a_{-2} & a_{-1} & a_0 & a_1 & \cdots \\
 & \vdots & \vdots & \vdots & \vdots & \vdots &
\end{array}\right\|,
$$

这里, $\Pr\{\xi=k\}=a_k, k=0,\pm1,\pm2,\cdots$, 且 $a_k\geqslant 0$, $\sum\limits_{k=-\infty}^{+\infty}a_k=1$.

B. 一维随机游动

在讨论随机游动时, 可把系统的状态直观地看作是一个移动粒子的位置.

一维随机游动是这样一类马尔可夫链, 其状态空间为整数集的有限或无限子集: $a,a+1,\cdots,b$, 如果某个粒子处于状态 i, 则通过一次转移它或者仍停留在 i, 或者移动到相邻状态 $i-1,i+1$ 其中之一. 如果状态空间是非负整数集, 则一个随机游动的转移矩阵有如下形式

48

$$
\boldsymbol{P}=\left\|\begin{array}{cccccccc}
r_0 & p_0 & 0 & 0 & \cdots & & & \\
q_1 & r_1 & p_1 & 0 & \cdots & & & \\
0 & q_2 & r_2 & p_2 & \cdots & & & \\
 & \ddots & & & & & & \\
 & & 0 & q_i & r_i & p_i & 0 & \\
 & & \ddots & & & & \ddots &
\end{array}\right\|, \tag{2.2}
$$

其中 $p_i>0, q_i>0, r_i\geqslant 0$, 并且 $q_i+r_i+p_i=1, i=1,2,\cdots(i\geqslant 1), p_0\geqslant 0, r_0\geqslant 0, r_0+p_0=1$. 特别地, 如果 $X_n=i$, 则对于 $i\geqslant 1$

$$
\Pr\{X_{n+1}=i+1|X_n=i\}=p_i,\quad \Pr\{X_{n+1}=i-1|X_n=i\}=q_i,
$$

$$
\Pr\{X_{n+1}=i|X_n=i\}=r_i,
$$

当 $i=0$ 时, 上面第一、三等式仍成立.

名称 "随机游动" 似乎是恰当的, 因为这个过程的一个直观描述是一个人 (如醉汉) 随机地向前或向后移动一步.

赌徒在一系列比赛中财富的变化通常用随机游动来描述. 明确地说, 假设一个拥有财富 k 的赌徒 A 和一个拥有无限财富的赌徒进行赌博, 赌徒 A 在每一次比赛中赢得一个单位财富的概率为 p_k, 输一个单位的概率为 $q_k = 1 - p_k, (k \geqslant 1)$, (每次比赛的选择可取决于他们的财富), 且设 $r_0 = 1$. 以 X_n 表示经过 n 次比赛之后赌徒 A 的财富, 过程 $\{X_n\}$ 显然是一个随机游动. 注意, 一旦达到状态 0(即赌徒 A 输光), 则过程将一直停留在这个状态. 这个过程通常也称为"赌徒输光"模型.

若对所有 $k \geqslant 1, p_k = p, q_k = 1 - p = q$, 且 $r_0 = 1$, 若 $p > q$ 则此随机游动描述在每次比赛中都有利于赌徒 A 的情形. 在第 3 章我们将证明赌徒 A 最终破产 (他输掉全部财富) 的概率为 $(q/p)^{x_0}$, 其中 x_0 表示在初始时刻 0 所拥有的财富, 而他的财富无限增加的概率是 $1 - (q/p)^{x_0}$. 若 $p < q$, 这时每次比赛有利于其对手, 赌徒 A 如果坚持赌下去, 则以概率 1 最终输光. 即使每次赌博都是公平的, 即 $p_k = q_k = \dfrac{1}{2}$, 结果也是一样的, 即赌徒 A 必定最终破产.

如果对手, 即赌徒 B, 起先也只有有限的财富 y, 而赌徒 A 拥有初始财富 x(设 $x + y = a$), 则我们现在重新考虑马尔可夫链 $\{X_n\}$, 这里 X_n 仍表示第 n 次比赛后赌徒 A 的财富. 然而现在过程的状态局限于值 $0, 1, 2, \cdots, a$. $a - X_n$ 表示第 n 次比赛后赌徒 B 的财富. 若在二次比赛中允许平局发生, 则该过程的转移概率矩阵为

49

$$\boldsymbol{P} = \left\| \begin{matrix} 1 & 0 & 0 & 0 & \cdots & \\ q_1 & r_1 & p_1 & 0 & \cdots & \\ 0 & q_2 & r_2 & p_2 & \cdots & \\ & & \ddots & & & \\ & & & q_{a-1} & r_{a-1} & p_{a-1} \\ 0 & \cdots & \cdots & 0 & 0 & 1 \end{matrix} \right\|, \tag{2.3}$$

其中 $p_i(q_i), i = 1, 2, \cdots, a - 1$, 表示当赌徒 A 的财富为 i 时接下来再增加 (减少) 一个单位财富的概率, 而 r_i 可以解释为发生平局的概率. 注意, 由 (2.3) 给出的马尔可夫链, 当赌徒 A 的财富 (或过程的状态) 达到 0 或 a 时, 以后就永远停留在此状态. 当过程到达状态 0 时, 我们说赌徒 A 输光了; 而当过程到达状态 a 时, 我们说赌徒 B 输光了.

随机游动不仅用来模拟赌博模型, 而且常用来离散逼近描述扩散粒子运动的物理过程. 如果一个粒子受到碰撞或随机脉冲, 它的位置是随机地移动的, 尽管粒子运动描述的是一连续轨道. 如果粒子将来的位置 (即它的概率分布) 仅仅取决于它现在的位置, 若以 X_t 表示时刻 t 的位置, 则过程 $\{X_t\}$ 是一个马尔可夫过程. 此连续运动的一个离散逼近对应着一个随机游动. 布朗运动 (见 1.2 节) 的古典离散形式是一个对称随机游动. 整数集上对称随机游动是这样一个马尔可夫链, 它的状态空

间是整数全体，转移概率矩阵的元素为

$$P_{ij}=\begin{cases}p, & 若\ j=i+1,\\ p, & 若\ j=i-1,\\ r, & 若\ j=i,\\ 0, & 其他.\end{cases}\qquad i,j=0,\pm1,\pm2,\cdots,$$

其中 $p>0, r\geqslant 0$，且 $2p+r=1$. 以下我们约定"对称随机游动"专指 $r=0, p=\frac{1}{2}$ 的情况.

50

为了考虑某种物理模型，我们需研究在非负整数集上的随机游动. 根据零状态的性质我们对不同的过程进行分类. 我们把注意力集中在由 (2.2) 所描述的随机游动上. 若 $p_0=1$，则 $r_0=0$. 这时零状态好像一个反射壁. 不管粒子什么时刻到达零状态，下一次转移必定回到状态 1. 这对应的物理过程相当于在零点存在一个弹性壁，粒子遇到弹性壁就无后效地反跳回来.

若 $p_0=0$ 且 $r_0=1$，则零相当于一个吸收壁，一旦粒子到达零，它就永远留在那里. 若 $p_0>0$ 且 $r_0>0$，则零是一个部分反射壁.

若随机游动限制在有限个状态 S，譬如 $0,1,2,\cdots,a$，则状态 0 和 a 都可以独立地或以任何组合方式成为反射壁、吸收壁或部分反射壁. 我们已经遇到一个限制在状态 S 的随机游动的模型 (例如，两个财产有限的赌徒输光模型)，其中 0 和 a 都是吸收的 [见 (2.3)].

透膜扩散的古典数学模型是有名的 Ehrenfest 模型，即边界状态为反射壁的有限状态集上的随机游动. 此随机游动的状态空间记为 $i=-a,-a+1,\cdots,-1,0,1,\cdots,a$，其转移概率矩阵元素为

$$P_{ij}=\begin{cases}\dfrac{a-i}{2a}, & 如果\ j=i+1,\\ \dfrac{a+i}{2a}, & 如果\ j=i-1,\\ 0, & 其他.\end{cases}$$

这个模型的物理解释如下. 想象两个共装着 $2a$ 个球的容器. 假设第一个容器 A 有 k 个球，第二个容器 B 有 $2a-k$ 个球. 从 $2a$ 个球中随机地挑选 1 个球 (即所有选择都是等可能的) 放到另一个容器. 每一次挑选都使过程发生一次转移. 易见，这些球是在两个容器之间波动，大体上是从装有较多球的容器移动到装有较少球的容器. 一个物理系统，若主要由正比于到达平衡状态的距离的恢复力所控制时，往往可以用 Ehrenfest 模型来近似.

古典的 n 维对称随机游动可以用下述方法描述. 状态空间是 n 维欧氏空间 E^n

51

的所有整数格点，即一个状态是由形如 $\boldsymbol{k}=(k_1,k_2,\cdots,k_n)$ 的一个 n 元整数组表示.

其转移概率矩阵的元素为

$$P_{kl}=\begin{cases}\dfrac{1}{2n}, & 若\displaystyle\sum_{i=1}^{n}|l_i-k_i|=1,\\ 0, & 其他.\end{cases}$$

类似于一维情况, E^n 上的对称随机游动表示 n 维布朗运动的一种离散形式.

C. 离散排队马尔可夫链

顾客到达服务台并排成队, 在每一段时间只有一个顾客受到服务假定此时至少有一个顾客出现. 如果没有顾客等候, 则在这段时间就没有人接受服务. (例如, 我们可想象一个出租汽车站, 在每个固定的时间段有一辆出租汽车到达并服务顾客, 若无人候车则出租汽车立即开走.) 在每一段服务时间内新顾客都有可能到达. 我们假设在第 n 段时间内实际到达的顾客数是一随机变量 ξ_n, 它的分布函数与时间段无关, 并由下式给出:

$$\begin{aligned}&\Pr\{在一段服务时间内有\ k\ 位顾客到达\}\\ =&\Pr\{\xi_n=k\}=a_k, k=0,1,\cdots,\\ &a_k\geqslant 0\quad 且\quad \sum_{k=0}^{\infty}a_k=1.\end{aligned}\tag{2.4}$$

我们还假设随机变量序列 ξ_n 是独立的. 在每段时间的开始系统的状态由排队等候服务的顾客数目确定. 如果现在的状态是 i, 那么经过这一段时间之后, 状态变为

$$j=\begin{cases}i-1+\xi, & 如果i\geqslant 1,\\ \xi, & 如果i=0,\end{cases}\tag{2.5}$$

这里 ξ 是在这一段时间内到达的新顾客数, 而在这段时间内只有一个顾客得到服务. 从随机过程的角度, 式 (2.5) 可表达为

$$X_{n+1}=(X_n-1)^{+}+\xi_n,\tag{2.6}$$

这里, $Y^{+}=\max(Y,0)$. 据式 (2.4) 和式 (2.5), 经过简单的计算, 其转移概率矩阵为

$$\|P_{ij}\|=\begin{Vmatrix}a_0 & a_1 & a_2 & a_3 & a_4 & \cdots\\ a_0 & a_1 & a_2 & a_3 & a_4 & \cdots\\ 0 & a_0 & a_1 & a_2 & a_3 & \cdots\\ 0 & 0 & a_0 & a_1 & a_2 & \cdots\\ 0 & 0 & 0 & a_0 & a_1 & \cdots\\ \vdots & & & & & \end{Vmatrix}.\tag{2.7}$$

显然, 如果在一段服务时间内到达的新顾客数目的期望值 $\sum_{k=0}^{\infty} ka_k$ 超过 1, 则排队队列长度随着时间推移将无限拉长.

另一方面, 如果 $\sum_{k=0}^{\infty} ka_k < 1$, 则我们将看到排队长度接近一个平衡状态 (稳定状态). 如果 $\sum ka_k = 1$, 则总体上是不稳定的. 这些叙述在我们建立了相关的常返理论 (见 3.5 节) 之后会得到详尽说明.

D. 存储模型

考虑商业上为了满足市场持续的需求而要在仓库准备一定的存货的情形. 我们假设货物的补充是发生在连续时刻 $t_1, t_2, \cdots$, 并假设在时间 (t_{n-1}, t_n) 的货物的累计需求量是一随机变量 ξ_n, 其分布函数与各段时间无关:

$$\Pr\{\xi_n = k\} = a_k, \qquad k = 0, 1, 2, \cdots, \tag{2.8}$$

这里 $a_k \geqslant 0$ 且 $\sum_{k=0}^{\infty} a_k = 1$ 在每段时间之初清点存货. 存储策略由两个事先指定的非负临界值 s 和 $S > s$ 来确定. 其具体执行办法如下: 倘若可用的存货不超过 s, 则马上进货以使现有存量达到 S. 要是可用存货已超过 s, 则不马上进货. 设 X_n 表示恰好在时刻 t_n 进货前的现有存储量. 过程 $\{X_n\}$ 的状态由所有可能存储值组成:

$$S, S-1, \cdots, +1, 0, -1, -2, \cdots,$$

其中, 负值可理解为超出库存的需求量, 因而需在再进货时直接供应. 依照存储原则, 相邻两个时间段的存储量有如下关系:

$$X_{n+1} = \begin{cases} X_n - \xi_{n+1}, & 若 s < X_n \leqslant S, \\ S - \xi_{n+1}, & 若 X_n \leqslant s, \end{cases} \tag{2.9}$$

这里 ξ_n 是在第 n 段时间里的需求量, 由分布律式 (2.8) 确定. 如果我们假设 ξ_n 是相互独立的, 那么存货量 $X_0, X_1, X_2, \cdots$ 显然构成一个马尔可夫链, 其转移概率矩阵可以由关系式 (2.9) 计算得到.

53

E. 成功的游程

考虑非负整数集上的一个马尔可夫链, 其转移概率矩阵为

$$\|P_{ij}\| = \begin{Vmatrix} p_0 & q_0 & 0 & 0 & \cdots \\ p_1 & 0 & q_1 & 0 & \cdots \\ p_2 & 0 & 0 & q_2 & \cdots \\ p_3 & 0 & 0 & 0 & \cdots \\ \vdots & & & & \end{Vmatrix}, \tag{2.10}$$

其中, $q_i > 0, p_i > 0$, 且 $q_i + p_i = 1, i = 0, 1, 2, \cdots$. 这里零状态起着特别的作用, 它可以从其他状态通过一次转移达到, 而状态 $i+1$ 只能由状态 i 转移达到.

这个例子便于计算, 因此我们将经常利用它解释一些概念和结果.

这个转移矩阵的一个特例是处理所谓"成功游程"问题得到的. 考虑一系列重复试验, 每次试验只有两种结果: 成功 (S) 和失败 (F). 并具体地说, 考虑只有两个结果 (S或F) 的一系列试验, 且假设在每一次试验中成功 (S) 的概率为 α, 而失败 (F) 的概率为 $\beta = 1-\alpha$. 如果到第 n 个试验时, 前 $r+1$ 次试验 (含第 n 次) 的试验结果分别为 $F, S, S, \cdots, S$, 则我们说在第 n 次试验出现了一个长度为 r 的成功游程. 现在我们把当前出现的成功游程的长度看作过程的当前状态. 特别地, 如果最后一次试验的结果是失败的, 则状态是零. 当最后 $r+1$ 次试验的结果依次为 $F, S, S, \cdots, S$, 则过程的状态变量将记为 r. 显然该过程是马尔可夫过程 (因为各次试验是彼此独立的), 并且其转移矩阵的形式为式 (2.10), 其中 $p_n = \beta, n = 0, 1, 2, \cdots$.

F. 分支过程

假定某生物体在其生命终结之际繁殖的后代数 ξ 是一随机变量, 其概率分布为

$$\Pr\{\xi = k\} = a_k, \quad k = 0, 1, 2, \cdots, \tag{2.11}$$

其中, $a_k \geqslant 0, \sum_{k=0}^{\infty} a_k = 1$. 我们假设所有这些后代的行动互相独立, 且在它们生命的最后按照概率分布式 (2.11) 各自繁殖后代 [为简单起见, 假定它们的寿命是相同的], 这样一代一代繁殖下去. 若 X_n 表示第 n 代的总数, 则过程 $\{X_n\}$ 是一个马尔可夫链.

54

事实上, 由于后代的数目仅是现有生物总数的函数, 因而关于 $X_{n_1}, X_{n_2}, \cdots, X_{n_r}, X_n, n_1 < n_2 < \cdots < n_r < n$ 的分布仅依赖于最后所知道的生物总数. 显然转移矩阵的元素由下式所给定:

$$P_{ij} = \Pr\{X_{n+1} = j | X_n = i\} = \Pr\{\xi_1 + \xi_2 + \cdots + \xi_i = j\}, \tag{2.12}$$

这里 $\xi_i (i = 1, 2, \cdots)$ 是对由分布律式 (2.11) 确定的随机变量的一次独立观测. 公式 (2.12) 可以简明地解释如下. 第 n 代的 i 个生物体各自独立繁殖它们后代的数目分别为 $\xi_1, \xi_2, \cdots, \xi_i$, 因此它们繁殖后代的累计数是 $\xi_1 + \xi_2 + \cdots + \xi_i$.

如果应用母函数, 显然 $\xi_1+\xi_2+\cdots+\xi_i$ 的母函数是 $[g(s)]^i$, 这里 g 是对应于 ξ 的母函数. (我们使用了独立随机变量和与母函数复合的性质, 见 1.1 节.) 因此 P_{ij} 就是 $[g(s)]^i$ 展成幂级数的第 j 个系数.

G. 遗传学中的马尔可夫链

下面理想化的遗传学模型是由 S.Wright 在研究基因频率受变异与选择影响而出现的波动现象时提出的. 我们首先来描述所谓单倍体的随机繁殖模型, 暂时不考虑变异现象和选择作用. 假设我们涉及的群体由总数为 $2N$ 的 a 型和 A 型基因构成. 则下一代的组成取决于如下 $2N$ 个独立二项试验: 如果原来群体由 j 个 a 型基因和 $2N-j$ 个 A 型基因组成, 则每次试验结果为 a 型或 A 型的概率分别为

$$p_j=\frac{j}{2N},\quad q_j=1-\frac{j}{2N}.$$

如此重复地选择. 依照这样的程序, 我们产生了一个马尔可夫链 $\{X_n\}$, 其中 X_n 是在第 n 代元素个数恒为 $2N$ 的群体当中 a 型基因的数目. 状态空间含有 $2N+1$ 个值 $\{0,1,2,\cdots,2N\}$, 转移概率矩阵依照二项分布来计算

55

$$\Pr\{X_{n+1}=k|X_n=j\}=P_{jk}=\begin{pmatrix}2N\\k\end{pmatrix}p_j^kq_j^{2N-k}(j,k=0,1,\cdots,2N). \qquad (2.13)$$

关于上述生物学原理方面的论述, 请读者参阅 Fisher 的书[1].

注意, 状态 0 和 $2N$ 是完全吸收的, 即当 $X_n=0$(或 $2N$) 时则对所有 $k\geqslant 0, X_{n+k}=0$(或 $2N$). 一个有趣的问题是确定在条件 $X_0=i$ 之下群体达到 "稳定", 即此群体纯粹由 a 型基因或 A 型基因组成的概率. 与此有关的是确定趋近于这个 "固定" 的速率. 我们将在第 3 章关于吸收概率的一般分析中, 研究这样的问题.

一个比较实际的模型是考虑变异过程. 我们假设在新一代形成之前, 每个基因有可能变异, 即变成另一种类型的基因. 特别地, 我们假设每个基因变异 $a\to A$ 出现的概率为 α_1, 变异 $A\to a$ 出现的概率为 α_2. 再者, 我们假设下一代的组成是由 $2N$ 个独立二项试验确定. 当母体群体由 j 个 a 型基因组成时, 二项试验的 p_j 和 q_j 的值取为

$$\begin{aligned}p_j&=\frac{j}{2N}(1-\alpha_1)+\left(1-\frac{j}{2N}\right)\alpha_2,\\ q_j&=\frac{j}{2N}\alpha_1+\left(1-\frac{j}{2N}\right)(1-\alpha_2).\end{aligned} \qquad (2.14)$$

其理由如下: 我们假设变异现象先发生, 然后新的基因由群体中随机选择确定. 这

1. R. A. Fisher, *The Genetical Theory of Natural Selection*, Oxford(Clarendon) Press, London and New York, 1962.

样, 在变异已经发生之后选择一个 a 型基因的概率恰好是当前 a 型基因数目的 $\frac{1}{2N}$, 因此平均概率 (关于可能的变异求平均) 显然是变异之后 a 型基因平均概率的 $\frac{1}{2N}$. 但是这个平均数正是 $j(1-\alpha_1)+(2N-j)\alpha_2$, 这就导出了式 (2.14).

与马尔可夫链相联系的转移概率由 (2.14) 所给出的 p_j, q_j 值以及式 (2.13) 来计算.

56

若 $\alpha_1, \alpha_2 > 0$, "稳定" 在任何状态都不会出现. 当 $n \to \infty$ 时, X_n 的分布函数将趋近于一个随机变量 ξ 的稳态分布, 其中 $\Pr\{\xi = k\} = \pi_k (k = 0, 1, 2, \cdots, 2N) \left(\sum_{k=0}^{2N} \pi_k = 1, \pi_k > 0\right)$. ξ 的分布函数称为稳态基因频率分布.

现在我们回头讨论简单随机配对模型并讨论有利于某基因 (例如 a 型基因) 的选择作用概率. 假设我们希望加强 a 型基因的选择优势, 使后代的相对数量的期望比例数分别为 $1+s$ 和 1, 这里 s 是较小的正数. 我们用

$$p_j = \frac{(1+s)j}{2N+sj}, \qquad q_j = 1 - p_j,$$

代替原来的 $p_j = j/2N$ 和 $q_j = 1 - j/2N$. 然后像前面一样通过二项抽样确定下一代基因. 如果母体群体由 j 个 a 型基因组成, 则下一代 a 型基因和 A 型基因的期望数分别是

$$2N \cdot \frac{(1+s)j}{2N+sj} \quad 和 \quad 2N \cdot \frac{2N-j}{2N+sj}.$$

第 $n+1$ 代 a 型基因和 A 型基因的期望数的比是

$$\frac{1+s}{1} \cdot \frac{j}{2N-j} = \frac{1+s}{1} \cdot \frac{第\ n\ 代中\ a\ 型基因的数目}{第\ n\ 代中\ A\ 型基因的数目}$$

由此说明了选择性的意义.

H. 遗传模型 II

基因表现为若干亚单位的组合, 为确定起见, 假设由 N 个亚单位组成, 当含有基因的细胞即将分裂时, 每个亚单位也分成两个, 两个细胞都接受了和前面相同数目亚单位构成的基因. 有一些亚单位可能处于变异的状态. 我们假定分裂时变异的单位产生变异的单位而非变异单位产生非变异单位. 再者, 两个新基因间亚单位的分配是随机的, 就像从一个罐中随机抽取这些单位. 我们只要跟踪遗传的一个单支而不必注意全部的繁殖情况. 为了描述某条单支的历史, 我们考虑一个马尔可夫链, 它的状态空间为 $\{0, 1, 2, \cdots, N\}$. 称一个基因处于状态 i, 如果它由 i 个变异亚单位

57

和 $N-i$ 个正常亚单位构成. 转移概率由下式计算:

$$P_{ij}=\binom{2i}{j}\binom{2N-2i}{N-j}\Big/\binom{2N}{N}. \tag{2.15}$$

此式源由如下: 假定母体基因处于状态 i, 分裂之后我们总共得到 $2i$ 个变异单位和 $2N-2i$ 个正常单位. 子体基因的状态由从这 $2N$ 个单位中随机挑出的 N 个单位构成, 根据超几何概率定律, 子体基因处于状态 j 的概率由 (2.15) 给出.

状态 $j=1,2,\cdots,N-1$ 称为混合的, 状态 0 和 N 称为单纯的. 状态 N 很有趣, 这时一个所有亚单位都是变异的基因, 可能引起该基因死亡, 而状态 0 说明这种类型的基因将不再发生变异. 我们以后将确定基因从状态 i 出发, 最终停留于状态 0 和 N 的具体概率.

2.3 马尔可夫链的转移概率矩阵

一个马尔可夫链完全由它的单步转移概率矩阵及在 0 时刻状态概率分布的详细描述所确定. 马尔可夫链的研究主要涉及过程可能实现概率的计算. 在这些计算中最重要的是 n 步转移概率矩阵 $\boldsymbol{P}^{(n)}=\|P_{ij}^n\|$. 这里 P_{ij}^n 表示过程经过 n 步转移从状态 i 到达状态 j 的概率, 即

$$P_{ij}^n=\Pr\{X_{n+m}=j|X_m=i\}. \tag{3.1}$$

这里应该注意, 我们暂时仅处理具有平稳转移概率的齐次过程, 否则, (3.1) 的左边还取决于 m.

由马尔可夫假设知, 我们可用一步转移概率矩阵 $\|P_{ij}\|$ 表达 (3.1) 定理叙述如下.

定理 3.1 如果一个马尔可夫链的单步转移概率矩阵是 $\boldsymbol{P}=\|P_{ij}\|$, 则对和为 n 的任意固定非负整数对 r,s 且 $r+s=n$,

58

$$P_{ij}^n=\sum_{k=0}^{\infty}P_{ik}^rP_{kj}^s, \tag{3.2}$$

这里, 我们规定

$$P_{ij}^0=\begin{cases}1, & i=j,\\ 0, & i\neq j.\end{cases}$$

由矩阵理论, 我们可看到关系式 (3.2) 恰好是矩阵乘积公式, 所以 $\boldsymbol{P}^{(n)}=\boldsymbol{P}^n$; 换句话说, P_{ij}^n 可以看成是 $\boldsymbol{P}$ 的 n 次幂矩阵 $\boldsymbol{P}^n$ 的元素.

证明 我们仅证明 $n=2$ 的情况, 经两步转移, 从状态 i 转移到状态 j 这个事件可以分解为若干事件, 首先从状态 i 第一步转移到中间状态 $k(k=0,1,2,\cdots)$, 然后再从状态 k 第二步转移到 j. 由马尔可夫假设, 第二步转移概率为 P_{kj}, 第一步转移概率为 P_{ik}. 利用全概率公式即导出式 (3.2). 一般情况的证明也是同样的. ■

如果过程初始状态为 j 的概率是 p_j, 即 X_0 的分布律是 $\Pr\{X_0=j\}=p_j$, 那么在时刻 n 过程处于状态 k 的概率为

$$p_k^{(n)}=\sum_{j=0}^{\infty}p_jP_{jk}^n=\Pr\{X_n=k\}. \tag{3.3}$$

除了确定过程的联合概率分布外, 一个比较困难但却是有趣的问题是研究当 $n\to\infty$ 时 P_{ij}^n 的渐近性质. 我们可以期望初始状态的影响随着时间推移而逐渐减小且当 $n\to\infty$ 时, P_{ij}^n 趋近一个与 i 无关的极限. 为了准确分析过程的渐近性质, 我们需要引入一些关于马尔可夫状态分类的原则.

2.4 马尔可夫链的状态分类

状态 j 称为**可自状态 i 到达的**, 如果对某个整数 $n\geqslant 0, P_{ij}^n>0$, 即状态 j 是可自状态 i 到达的, 如果从状态 i 出发经过有限步转移到达状态 j 的概率是正的. 两个状态 i 和 j 彼此可达, 则称为**互通的**, 并记为 $i\longleftrightarrow j$. 如果状态 i 和 j 不是互通的, 则关系式 [59]

$$P_{ij}^n=0 \quad 对所有n\geqslant 0$$

和

$$P_{ji}^n=0 \quad 对所有n\geqslant 0$$

至少成立一个. 互通的概念是一等价关系:

(i) $i\longleftrightarrow i$(自反性), 这可由定义 $P_{ij}^0=\delta_{ij}=\left\{\begin{array}{ll}1, & i=j\\ 0, & i\neq j\end{array}\right\}$ 直接推得.

(ii) 若 $i\longleftrightarrow j$, 则 $j\longleftrightarrow i$(对称性) 这由互通的定义推得.

(iii) 若 $i\longleftrightarrow j$ 且 $j\longleftrightarrow k$, 则 $i\longleftrightarrow k$(传递性).

关于传递性, 证明如下: $i\longleftrightarrow j$ 和 $j\longleftrightarrow k$ 蕴含着存在整数 n 和 m 使得 $P_{ij}^n>0$ 和 $P_{jk}^m>0$. 因此由式 (3.2) 和每个 P_{rs}^t 的非负性, 推得

$$P_{ik}^{n+m}=\sum_{r=0}^{\infty}P_{ir}^nP_{rk}^m\geqslant P_{ij}^nP_{jk}^m>0.$$

类似可证, 存在正整数 v 使得 $P_{ki}^v>0$.

现在我们可把过程状态的全体分成若干个等价类. 在同一个等价类里每个状态都是互通的. 可能会出现从某一类出发以正概率转移到另一类, 但相反方向的转移就不可能了, 否则这两类就归并成一个等价类. 我们称马尔可夫链是**不可约的**, 如果等价关系只约化出一个类, 即一个过程是不可约的, 如果所有状态彼此互通.

为了解释这个概念, 我们考虑转移概率矩阵

$$P = \left\| \begin{array}{cc|ccc} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 \end{array} \right\| = \left\| \begin{array}{cc} P_1 & 0 \\ 0 & P_2 \end{array} \right\|,$$

这里 P_1 是矩阵 P 前两行和前两列交叉的元素构成的矩阵, 矩阵 P_2 是由矩阵 P 后 [60] 三行三列交叉元素构成的. 这个马尔可夫链的状态显然分成两类: $\{1,2\}$ 和 $\{3,4,5\}$. 如果 X_0 的状态处于第一类, 则该系统的状态此后一直停留在这一类中, 其相应的转移矩阵是 P_1. 类似地, 如果初始状态属于第二类, 其相应的转移矩阵是 P_2. 这相当于我们把两个完全无关的过程组合在一起.

在随机游动模型中, 若其转移概率矩阵为

$$P = \left\| \begin{array}{cccccccc} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ q & 0 & p & 0 & \cdots & 0 & 0 & 0 \\ 0 & q & 0 & p & \cdots & 0 & 0 & 0 \\ \vdots & & & & & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & & & q & 0 & p \\ 0 & \cdots & \cdots & & & 0 & 0 & 1 \end{array} \right\| \begin{array}{c} \text{状态} \\ 0 \\ 1 \\ 2 \\ \vdots \\ a-1 \\ a \end{array},$$

则此时我们有三个等价类: $\{0\},\{1,2,\cdots,a-1\}$ 和 $\{a\}$. 这个例子表明可以从第二类出发到达第一类或第三类, 但相反方向的转移是不可能的.

可直接验证, 当对于所有 k 都有 $a_k > 0$ 时, 排队马尔可夫链 (2.2 节例 C) 是不可约的. 在相同条件下易证存储模型 (例 D) 也是不可约的. 成功游程的马尔可夫链模型 (例 F) 在条件 $q_i > 0, p_i > 0 (i = 0, 1, \cdots)$ 下同样是不可约的.

马尔可夫链的周期性

对于某个 i, 使 $P_{ii}^n > 0$ 成立的所有正整数 n 的最大公约数 (*g.c.d.*) 称为状态 i 的周期, 记为 $d(i)$. [如果对所有 $n \geqslant 1, P_{ii}^n = 0$, 则定义 $d(i) = 0$]. 如果在随机游动

(见式 (2.2)) 中所有 $r_i = 0$, 则每个状态的周期为 2. 如果对某个状态 $i_0, r_{i0} > 0$, 那么每个状态的周期为 1, 因为这时不论初始状态 j 如何, 系统都能到达状态 i_0, 并且在回到状态 j 之前停留在状态 i_0 的时间可任意长.

具有 n 个状态的有限马尔可夫链, 若其转移矩阵为

$$P = \left\| \overbrace{\begin{matrix} 0 & 1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & & & & & & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & \cdots & 1 \\ 1 & 0 & 0 & \cdots & \cdots & \cdots & 0 \end{matrix}}^{n} \right\|$$

61

则每个状态的周期都是 n.

我们列出状态周期的三个基本性质但不给予证明 (这方面问题可参阅本章问题 2-4).

定理 4.1 如果 $i \longleftrightarrow j$, 则 $d(i) = d(j)$.

此结论表明在每个互通的等价类中周期是一常数.

定理 4.2 若状态 i 具有周期 $d(i)$, 则存在一个与 i 有关的整数 N, 使得对所有 $n \geqslant N$

$$P_{ii}^{nd(i)} > 0.$$

此结论说明从状态 i 出发, 经过周期的足够大信数时间后, 回到状态 i 的概率均大于 0.

推论 4.1 若 $P_{ji}^m > 0$, 则 $P_{ji}^{m+nd(i)} > 0$ 对所有足够大的正整数 n 都成立.

一个马尔可夫链, 若其每个状态的周期都为 1, 则称为**非周期的**. 我们遇见的绝大多数马尔可夫链过程都是非周期的. 随机游动通常作为实际问题中周期情形的典型代表. 后面将介绍非周期情况的一些结果, 对一般情况, 相应的结论通常只叙述而不加证明, 相信勤奋的读者能够很容易地把证明补上.

2.5 常 返 性

考虑任意一个固定的状态 i. 对每个整数 $n \geqslant 1$, 我们定义

$$f_{ii}^n = \Pr\{X_n = i, X_v \neq i, v = 1, 2, \cdots, n-1 | X_0 = i\}.$$

换句话说, f_{ii}^n 是从状态 i 出发, 在第 n 次转移时首次回到状态 i 的概率. 显然 $f_{ii}^1 = P_{ii}$ 且 f_{ii}^n 可依照如下公式递推计算:

$$P_{ii}^{n}=\sum_{k=0}^{n}f_{ii}^{k}P_{ii}^{n-k},\quad n\geqslant 1,\tag{5.1}$$

其中对于所有的 i 我们定义 $f_{ii}^{0}=0$. 等式 (5.1) 是把计算 P_{ii}^{n} 的事件按照首返状态 i 的时间进行分解而导出的. 事实上, 我们来考虑这个过程在已知 $X_0=i$ 的条件下 $X_n=i$ 且在 k 次转移后首次回到状态 i 的所有可能的实现, 记此事件为 E_k. 显然事件 $E_k(k=1,2,\cdots,n)$ 是互斥的. 从 $X_0=i$ 出发, 经过 k 次转移首次回到状态 i 的概率定义为 f_{ii}^{k}, 在余下的 $n-k$ 次转移中我们只关心那些使 $X_n=i$ 的实现. 利用马尔可夫性质,

62

$$\begin{aligned}\Pr\{E_k\}=\Pr\{&\text{在 } k \text{ 次转移时首次返回}|X_0=i\}\cdot\\&\Pr\{X_n=i|X_k=i\}=f_{ii}^{k}\cdot P_{ii}^{n-k},1\leqslant k\leqslant n,\end{aligned}$$

(注意, $P_{ii}^{0}=1$). 因此

$$\begin{aligned}\Pr\{X_n=i|X_0=i\}&=\sum_{k=1}^{n}\Pr\{E_k\}\\&=\sum_{k=1}^{n}f_{ii}^{k}P_{ii}^{n-k}=\sum_{k=0}^{n}f_{ii}^{k}P_{ii}^{n-k},\end{aligned}$$

最后的等号根据定义 $f_{ii}^{0}=0$.

下面我们介绍相关的母函数.

定义 序列 $\{P_{ij}^{n}\}$ 的母函数 $P_{ij}(s)$ 是

$$P_{ij}(s)=\sum_{n=0}^{\infty}P_{ij}^{n}s^{n},\quad |s|<1.\tag{5.2}$$

用类似方法也可定义序列 $\{f_{ij}^{n}\}$(注: 当 $i\neq j$ 时 f_{ij}^{n} 的定义见下面等式 (5.9)) 的母函数是

$$F_{ij}(s)=\sum_{n=0}^{\infty}f_{ij}^{n}s^{n},\quad |s|<1.\tag{5.3}$$

回顾如下事实[1]: 如果

$$A(s)=\sum_{k=0}^{\infty}a_k s^k\quad \text{且}\quad B(s)=\sum_{l=0}^{\infty}b_l s^l,\tag{5.4}$$

1.
$$\begin{aligned}A(s)B(s)&=\left(\sum_{k=0}^{\infty}a_k s^k\right)\left(\sum_{l=0}^{\infty}b_l s^l\right)\\&=a_0b_0+(a_1b_0+b_1a_0)s+(a_2b_0+a_1b_1+a_0b_2)s^2+\cdots\\&=\sum_{k=0}^{\infty}s^k\left(\sum_{j=0}^{k}a_jb_{k-j}\right)=\sum_{k=0}^{\infty}c_k s^k.\end{aligned}$$

则

$$A(s)B(s) = C(s) = \sum_{r=0}^{\infty} c_r s^r, \quad |s| < 1, \tag{5.5}$$

其中

$$c_r = a_0 b_r + a_1 b_{r-1} + \cdots + a_r b_0. \tag{5.6}$$

如果我们把 f_{ii}^k 看作 a_k, P_{ii}^l 看作 b_l, 则比较式 (5.1) 和式 (5.6) 即得到

$$F_{ii}(s)P_{ii}(s) = P_{ii}(s) - 1, \quad |s| < 1, \tag{5.7}$$

或

$$P_{ii}(s) = \frac{1}{1 - F_{ii}(s)}, \quad |s| < 1. \tag{5.8}$$

在式 (5.7) 中减 1 是必要的, 因为当 $n = 0$ 时式 (5.1) 不成立.

利用证明式 (5.1) 的类似方法可得到

$$P_{ij}^n = \sum_{k=0}^{n} f_{ij}^k P_{jj}^{n-k}, \quad i \neq j, n \geqslant 0, \tag{5.9}$$

其中 f_{ij}^k 是从状态 i 出发经过 k 次转移首次达到 j 的概率. 对所有 i 和 j, 我们重新定义 $f_{ij}^0 = 0$. 对照式 (5.5), 由式 (5.9) 推得

$$P_{ij}(s) = F_{ij}(s)P_{jj}(s), \quad |s| < 1. \tag{5.10}$$

我们说状态 i 是**常返的**, 当且仅当 $\sum_{n=1}^{\infty} f_{ii}^n = 1$. 这就是说, 状态 i 是常返的. 当且仅当从状态 i 出发经过有限时长之后回到状态 i 的概率为 1, 非常返的状态又称为**瞬态**. 下面我们来证明常返的或非常返的状态与 n 步转移概率 P_{ii}^n 关系的一个定理. 在证明之前, 需要如下引理:

引理 5.1 (阿贝尔)

(a) 如果 $\sum_{k=0}^{\infty} a_k$ 收敛, 则

$$\lim_{s \to 1^-} \sum_{k=0}^{\infty} a_k s^k = \sum_{k=0}^{\infty} a_k = a \tag{5.11}$$

($\lim_{s \to 1^-}$ 指 s 从小于 1 的方向趋于 1).

(b) 如果 $a_k \geqslant 0$ 且 $\lim_{s \to 1^-} \sum_{k=0}^{\infty} a_k s^k = a \leqslant \infty$, 则

$$\sum_{k=0}^{\infty} a_k = \lim_{N \to \infty} \sum_{k=0}^{N} a_k = a.$$

证明　(a) 我们要证明

$$\lim_{s\to 1^-}\left|\sum_{k=0}^{\infty}a_k(s^k-1)\right|=0. \tag{5.12}$$

因为 $\sum_{k=0}^{\infty}a_k$ 收敛, 对任意 $\varepsilon>0$, 我们可找到这样一个 $N(\varepsilon)$ 使得对所有 $N'\geqslant N$ 有

$$\left|\sum_{k=N}^{N'}a_k\right|<\frac{\varepsilon}{4}.$$

选取这样的 N 后,

64

$$\begin{aligned}\left|\sum_{k=0}^{\infty}a_k(s^k-1)\right|&=\left|\sum_{k=0}^{N}a_k(s^k-1)+\sum_{k=N+1}^{\infty}a_k(s^k-1)\right|\\&\leqslant\left|\sum_{k=0}^{N}a_k(s^k-1)\right|+\left|\sum_{k=N+1}^{\infty}a_k(s^k-1)\right|.\end{aligned} \tag{5.13}$$

对于 $0\leqslant s<1$ 有,

$$\left|\sum_{k=0}^{N}a_k(s^k-1)\right|\leqslant MN\left|s^N-1\right|, \tag{5.14}$$

其中 $M=\max\limits_{0\leqslant k\leqslant N}|a_k|<\infty$, 所以对充分靠近于 1 的 s, 我们有

$$\left|\sum_{k=0}^{N}a_k(s^k-1)\right|<\frac{\varepsilon}{2}.$$

为了估计 $\sum_{k=N+1}^{\infty}a_k(s^k-1)$, 我们分部求和, 记 $A_k=\sum_{r=k}^{\infty}a_r$, 于是

$$\begin{aligned}\left|\sum_{k=N+1}^{\infty}a_k(s^k-1)\right|&=\left|\sum_{k=N+1}^{\infty}(A_k-A_{k+1})(s^k-1)\right|\\&=\left|A_{N+1}(s^{N+1}-1)+\sum_{k=N+2}^{\infty}A_k(s^k-s^{k-1})\right|,\end{aligned} \tag{5.15}$$

显然, 式 (5.15) 右边不超过

$$\frac{\varepsilon}{4}\left|(s^{N+1}-1)\right|+\frac{\varepsilon}{4}s^{N+1}\leqslant\frac{\varepsilon}{2}.$$

综上所述, 当 s 充分靠近于 1 时, 我们有

$$\left|\sum_{k=0}^{\infty} a_k(s^k - 1)\right| < \varepsilon.$$ ■

(b) 对于 $0 < s < 1, \sum_{k=0}^{\infty} a_k s^k \leqslant \sum_{k=0}^{\infty} a_k$. 所以当 $a = \infty$ 时是显然的. 若 $a < \infty$, 则由我们的假设

$$\sum_{k=0}^{\infty} a_k s^k < a < \infty \quad 当 0 < s < 1.$$

因此, 对所有 n

$$\sum_{k=0}^{n} a_k \leqslant a.$$

由于 $\sum_{k=0}^{n} a_k$ 是 n 的有界单调增函数, 因此它存在有限极限, 记为 a'. 但由 (a) 立即推得 $a' = a$. ■

65

由此引理我们容易证明

定理 5.1 状态 i 是常返的, 当且仅当

$$\sum_{n=1}^{\infty} P_{ii}^n = \infty.$$

证明 假设 i 是常返的, 即 $\sum_{n=1}^{\infty} f_{ii}^n = 1$. 则由引理 5.1(a)

$$\lim_{s\to 1^-} \sum_{n=0}^{\infty} f_{ii}^n s^n = \lim_{s\to 1^-} F_{ii}(s) = 1.$$

这样, 从式 (5.8) 得

$$\lim_{s\to 1^-} P_{ii}(s) = \lim_{s\to 1^-} \sum_{n=0}^{\infty} P_{ii}^n s^n = \infty.$$

由引理 5.1(b), 我们有

$$\sum_{n=0}^{\infty} P_{ii}^n = \infty,$$

必要性得证. 为了证明充分性, 假设状态 i 是瞬态的, 即 $\sum_{n=1}^{\infty} f_{ii}^n < 1$. 利用引理 5.1(a) 并结合式 (5.8), 我们推得

$$\lim_{s\to 1^-} P_{ii}(s) < \infty.$$

现在再根据引理 5.1(b), 我们得到 $\sum_{n=1}^{\infty} P_{ii}^n < \infty$, 这与假设矛盾, 从而充分性得证. ■

由定理 5.1 我们立刻得到一个推论.

推论 5.1 如果 $i \longleftrightarrow j$ 且 i 是常返的, 则 j 也是常返的.

证明 由于 $i \longleftrightarrow j$, 则存在 m, n, 使得

$$P_{ij}^n > 0 \quad 和 \quad P_{ji}^m > 0.$$

设 $v > 0$. 利用通常证法 (见 2.4 节) 可得 $P_{jj}^{m+n+v} \geqslant P_{ji}^m P_{ii}^v P_{ij}^n$, 然后求和, 得

$$\sum_{v=0}^{\infty} P_{jj}^{m+n+v} \geqslant \sum_{v=0}^{\infty} P_{ji}^m P_{ii}^v P_{ij}^n = P_{ji}^m P_{ij}^n \sum_{v=0}^{\infty} P_{ii}^v.$$

因此, 若 $\sum_{v=0}^{\infty} P_{ii}^v$ 发散, 则 $\sum_{v=0}^{\infty} P_{jj}^v$ 也发散. ■

这个推论表明, 常返性也像周期性一样是一个类性质: 即在一个等价类中的状态或者都是常返的, 或者都不是.

附注 给定 $X_0 = i$, 返回状态 i 的次数的期望是 $\sum_{n=1}^{\infty} P_{ii}^n$. 这样, 定理 5.1 表明, 状态 i 为常返的当且仅当返回次数的期望值无限.

66

2.6 常返马尔可夫链的例子

例 1 首先考虑在正整数和负整数上的一维随机游动, 在每次转移中粒子向右移动一个单位的概率为 p, 向左移动一个单位的概率为 $q(p+q=1)$. 因此

$$P_{00}^{2n+1} = 0, \quad n = 0, 1, 2, \cdots,$$

$$P_{00}^{2n} = \binom{2n}{n} p^n q^n = \frac{(2n)!}{n!n!} p^n q^n. \tag{6.1}$$

借助于斯特林公式,

$$n! \sim n^{n+\frac{1}{2}} \mathrm{e}^{-n} \sqrt{2\pi}. \tag{6.2}$$

应用式 (6.2) 于式 (6.1), 我们得到

$$P_{00}^{2n} \sim \frac{(pq)^n 2^{2n}}{\sqrt{\pi n}} = \frac{(4pq)^n}{\sqrt{\pi n}}.$$

容易验证 $p(1-p)=pq\leqslant\frac{1}{4}$, 当且仅当 $p=q=\frac{1}{2}$ 时等号成立. 因此, 当且仅当 $p=\frac{1}{2}$ 时 $\sum_{n=0}^{\infty}P_{00}^{n}=\infty$. 故而, 由定理 5.1 知, 当且仅当 $p=q=\frac{1}{2}$ 时一维随机游动是常返的. 记住, 常返性是一个类性质. 可以想象, 如果 $p>q$, 初始位于原点的粒子趋向于 $+\infty$(若 $p<q$, 趋向于 $-\infty$) 而不再回到原点的概率是正的.

例 2 现在我们来看一下在全平面上的二维随机游动. 设从某个状态往上、下、左、右移动一个单位的概率都等于 1/4. 我们仍然研究由原点代表的状态的常返性. 考虑向右移动 i 个单位、向左移动 i 个单位、向下移动 j 个单位、向上移动 j 个单位的所有可能的路线, 其中 $2i+2j=2n$. 通过计算我们有

$$\begin{aligned}&P_{00}^{2n+1}=0,\qquad n=0,1,2,\cdots,\\&P_{00}^{2n}=\sum_{i,j,i+j=n}\frac{(2n)!}{i!i!j!j!}\left(\frac{1}{4}\right)^{2n},\qquad n=1,2,3,\cdots.\end{aligned}\tag{6.3}$$

[式 (6.3) 是依据多项分布推得.] 式 (6.3) 的分子和分母同乘以 $(n!)^2$, 得到

$$P_{00}^{2n}=\left(\frac{1}{4}\right)^{2n}\binom{2n}{n}\sum_{i=0}^{n}\binom{n}{i}\binom{n}{n-i}.$$

67

但

$$\sum_{i=0}^{n}\binom{n}{i}\binom{n}{n-i}=\binom{2n}{n}.$$

因此

$$P_{00}^{2n}=\left(\frac{1}{4}\right)^{2n}\binom{2n}{n}^2.$$

再一次利用斯特林公式 (6.2), 我们得到

$$P_{00}^{2n}\sim\frac{1}{\pi n}.$$

因此, $\sum_{n=0}^{\infty}P_{00}^{n}=\infty$, 即由原点代表的状态也是常返状态.

例 3 我们考虑在三维空间里的对称随机游动. 同前面一样, 我们得到

$$\begin{aligned}&P_{00}^{2n+1}=0,\qquad n=0,1,2,\cdots,\\&P_{00}^{2n}=\sum_{i,j,0\leqslant i+j\leqslant n}\frac{(2n)!}{i!i!j!j!(n-i-j)!(n-i-j)!}\left(\frac{1}{6}\right)^{2n}.\end{aligned}\tag{6.4}$$

用 $(n!)^2$ 乘以分子分母并提出因子 $\left(\frac{1}{2}\right)^{2n}$, 得

$$P_{00}^{2n}=\frac{1}{2^{2n}}\binom{2n}{n}\sum_{i,j,0\leqslant i+j\leqslant n}\left[\frac{n!}{i!j!(n-i-j)!}\right]^2\left(\frac{1}{3}\right)^{2n}, \tag{6.5}$$

$$P_{00}^{2n}\leqslant c_n\frac{1}{2^{2n}}\binom{2n}{n}\frac{1}{3^n}, \tag{6.6}$$

其中

$$c_n=\max_{i,j,0\leqslant i+j\leqslant n}\left[\frac{n!}{i!j!(n-i-j)!}\right]. \tag{6.7}$$

注意, 上面我们利用了等式

$$\sum_{i,j,0\leqslant i+j\leqslant n}\frac{n!}{i!j!(n-i-j)!}\left(\frac{1}{3}\right)^n=1. \tag{6.8}$$

对于较大的 n, 当 $i=j\sim\frac{n}{3}$ 时, c_n 值可以达到. 下面我们来证明这个结果. 设 i_0 和 j_0 使得

68

$$\frac{n!}{i!j!(n-i-j)!}\qquad(0\leqslant i+j\leqslant n).$$

达到最大值. 下面四个不等式是显然的:

$$\begin{aligned}
\frac{n!}{j_0!(i_0-1)!(n-j_0-i_0+1)!}&\leqslant\frac{n!}{j_0!i_0!(n-j_0-i_0)!},\\
\frac{n!}{j_0!(i_0+1)!(n-j_0-i_0-1)!}&\leqslant\frac{n!}{j_0!i_0!(n-j_0-i_0)!},\\
\frac{n!}{(j_0-1)!i_0!(n-j_0-i_0+1)!}&\leqslant\frac{n!}{j_0!i_0!(n-j_0-i_0)!},\\
\frac{n!}{(j_0+1)!i_0!(n_0-j_0-i_0-1)!}&\leqslant\frac{n!}{j_0!i_0!(n-j_0-i_0)!}.
\end{aligned}$$

这些不等式简化为

$$\begin{aligned}
n-i_0-1\leqslant 2j_0\leqslant n-i_0+1,\\
n-j_0-1\leqslant 2i_0\leqslant n-j_0+1.
\end{aligned}$$

因此, 对较大的 $n, i_0\sim n/3, j_0\sim n/3$. 把 $i=j=\frac{n}{3}$ 代入式 (6.7), 式 (6.6) 变为

$$P_{00}^{2n}\leqslant\frac{n!}{(n/3)!(n/3)!(n/3)!(2)^{2n}3^n}\binom{2n}{n}. \tag{6.9}$$

若我们利用斯特林公式, 不等式 (6.9) 右边渐近于

$$\frac{3\sqrt{3}}{2\pi^{3/2}n^{3/2}}.$$

对上面这些项求和, 得

$$\sum_{n=1}^{\infty}\frac{3\sqrt{3}}{2\pi^{3/2}n^{3/2}}<\infty.$$

因此 $\sum\limits_{n=1}^{\infty}P_{00}^{n}<\infty$, 由定理 5.1, 0 所表示的状态是瞬态. 由于常返性是一个类性质且所有状态互通, 我们已看到在一维和二维对称随机游动中粒子必然能返回原来所处的任何一个状态, 然而, 在三维对称随机游动中, 粒子一旦离开某个状态之后, 不再返回该状态的概率却是正的.

69

例 4 现在考虑描述二项试验成功游程的马尔可夫链, 其转移概率矩阵是

$$\left\|\begin{array}{ccccccc}
p_0 & 1-p_0 & 0 & 0 & \cdots & \cdots & \cdots \\
p_1 & 0 & 1-p_1 & 0 & \cdots & \cdots & \cdots \\
p_2 & 0 & 0 & 1-p_2 & \cdots & \cdots & \cdots \\
\vdots & \vdots & & & & & \\
p_r & 0 & \cdots & \cdots & 1-p_r & 0 & \cdots \\
\vdots & & & & \vdots & \vdots &
\end{array}\right\| \quad (0<p_i<1).$$

这个马尔可夫链的状态全部属于同一个等价类 (从任一状态出发都可以达到其他任何状态). 因为常返性是一个类性质 (见推论 5.1), 所以我们仅研究状态 0 的常返性. 经计算得

$$\begin{aligned}
f_{00}^{1}&=p_0=1-(1-p_0),\\
f_{00}^{n}&=\left(\prod_{i=0}^{n-2}(1-p_i)\right)p_{n-1},\quad n>1.
\end{aligned}\tag{6.10}$$

式 (6.10) 可重写为

$$f_{00}^{n}=\prod_{i=0}^{n-2}(1-p_i)[1-(1-p_{n-1})],\quad n>1,$$

$$f_{00}^{n}=(1-p_0)(1-p_1)\cdots(1-p_{n-2})-(1-p_0)(1-p_1)\cdots(1-p_{n-1}),\quad n>1,$$

记

$$u_n=\begin{cases}(1-p_0)(1-p_1)\cdots(1-p_n), & n\geqslant 0,\\ 1, & n=-1.\end{cases}$$

对 $f_{00}^n(n \geqslant 1)$ 求和, 我们有

$$\sum_{n=1}^{m+1} f_{00}^n = \sum_{n=1}^{m+1}(u_{n-2} - u_{n-1})$$
$$= (1-u_0) + (u_0 - u_1) + \cdots + (u_{m-1} - u_m)$$

或

$$\sum_{n=1}^{m+1} f_{00}^n = 1 - u_m. \tag{6.11}$$

下面的论证我们需要如下引理:

引理 6.1 如果 $0 < p_i < 1, i = 0,1,2,\cdots$, 则 $u_m = \prod_{i=0}^{m}(1-p_i) \to 0 (m \to \infty)$ 的充分必要条件是 $\sum_{i=0}^{\infty} p_i = \infty$.

证明 设 $\sum_{i=0}^{\infty} p_i = \infty$. 由于 $\exp(-p_i)$ 的级数展开是一交错级数, 且其各项绝对值是递减的, 我们得

70

$$1 - p_i < 1 - p_i + \frac{p_i^2}{2!} - \frac{p_i^3}{3!} + \cdots = \exp(-p_i), i = 0,1,2,\cdots. \tag{6.12}$$

由于 (6.12) 对所有 i 成立, 从而 $\prod_{i=0}^{m}(1-p_i) < \exp\left(-\sum_{i=0}^{m} p_i\right)$. 但是由假设, $\lim_{m\to\infty} \sum_{i=0}^{m} p_i = \infty$, 因此 $\lim_{m\to\infty} \prod_{i=0}^{m}(1-p_i) = 0$. 为证必要性, 直接由归纳法易知, 对任意 j 及所有 $m = j+1, j+2, \cdots$,

$$\prod_{i=j}^{m}(1-p_i) > (1 - p_j - p_{j+1} - \cdots - p_m).$$

现假设 $\sum_{i=1}^{\infty} p_i < \infty$, 则对某个 $j > 1$, 成立 $0 < \sum_{i=j}^{\infty} p_i < 1$. 这样,

$$\lim_{m\to\infty} \prod_{i=j}^{m}(1-p_i) > \lim_{m\to\infty}(1 - \sum_{i=j}^{m} p_i) > 0,$$

这与 $u_m \to 0$ 矛盾. ■

现在我们回到式 (6.11), 应用引理 6.1 立刻推得: $\sum_{n=1}^{\infty} f_{00}^n = 1$ 当且仅当 $\sum_{i=0}^{\infty} p_i = \infty$, 即状态 0 是常返的当且仅当 p_i 之和发散.

我们顺便指出如下事实. 给定任意一集合 $\{a_1, a_2, \cdots\}, a_i > 0$, 且 $\sum_{i=1}^{\infty} a_i \leqslant 1$, 我们可以找到一列 $p_i, 0 < p_i < 1$, 由它所确定的上述马尔可夫链有 $f_{00}^n = a_n$. 为此, 令

$$f_{00}^1 = p_0 = a_1,$$
$$f_{00}^2 = (1 - p_0)p_1 = a_2,$$

由此确定

$$p_1 = \frac{a_2}{1 - a_1}.$$

记

$$f_{00}^3 = (1 - p_0)(1 - p_1)p_2 = a_3,$$

这蕴含着

$$p_2 = \frac{a_3}{1 - a_1 - a_2}.$$

如此类推, 我们得到满足 $0 < p_i < 1$ 的一列 p_i.

71

2.7 关于常返性的补充

下面定理表明, 若一个指定状态一定是常返的, 则此状态将以概率 1 出现无限次. 记

$Q_{ii} = \Pr\{$ 一个粒子从状态 i 出发无限次地返回状态 $i\}$.

定理 7.1 状态 i 是常返的或瞬时的, 分别对应于 Q_{ii} 等于 1 或 0.

证明 设 Q_{ii}^N 定义为

$Q_{ii}^N = \Pr\{$一个粒子从状态 i 出发至少 N 次返回状态 $i\}$.

按照首次返回时间来分解事件 Q_{ii}^N, 我们有等式

$$Q_{ii}^N = \sum_{k=1}^{\infty} f_{ii}^k Q_{ii}^{N-1} = Q_{ii}^{N-1} f_{ii}^*,$$

其中, $f_{ii}^* = \sum_{k=1}^{\infty} f_{ii}^k$.

递推, 得到

$$Q_{ii}^N = f_{ii}^* Q_{ii}^{N-1} = (f_{ii}^*)^2 Q_{ii}^{N-2} = \cdots = [f_{ii}^*]^{N-1} Q_{ii}^1.$$

但, 由定义 $Q_{ii}^1 = f_{ii}^*$. 因此

$$Q_{ii}^N = [f_{ii}^*]^N.$$

由于 $\lim_{N\to\infty} Q_{ii}^N = Q_{ii}$, 我们得到 $Q_{ii}=1$ 或 0 分别对应于 $f_{ii}^*=1$ 或 <1, 即对应于状态 i 是常返的或瞬时的.

定理 7.2 若 $i \longleftrightarrow j$ 且类是常返的, 则

$$f_{ij}^* = \sum_{n=1}^{\infty} f_{ij}^n = 1.$$

证明简单, 从略.

我们定义符号 Q_{ij} 为

Pr{ 粒子从状态 i 出发无限次到达状态 j}.

由定理 7.2 立即得到如下推论.

推论 7.1 如果 $i \longleftrightarrow j$ 且类是常返的, 那么 $Q_{ij}=1$.

证明 易知

72

$$Q_{ij} = f_{ij}^* Q_{jj}.$$

即然 j 是常返状态, 由定理 7.1 知, $Q_{jj}=1$. 再由定理 7.2 知, $f_{ij}^*=1$. 因此 $Q_{ij}=1$. ■

初等问题

1. 试求下列马尔可夫链的转移概率矩阵 $\|P_{jk}\|$:

(a) 考虑一组抛掷硬币的试验, 每次试验出现正面的概率为 p, 在时刻 n(即在 n 次抛掷后) 过程的状态是 n 次抛掷中出现的正面数减去出现的反面数.

(b) N 个黑球和 N 个白球放在两个罐里, 使得每个罐装有 N 个球. 每次从两个罐中各随机抽取一个球, 相互交换后又放回罐中. 系统的状态是第一个罐里的白球数.

(c) 一只白鼠放进如下图所示的迷宫, 白鼠随机地从一个格子移动到相邻格子, 即, 如果它离开一个格子的途径有 k 种, 则选择其中一条途径的概率为 $1/k$. 这只白鼠每单位时间改变一个格子. 系统的状态是白鼠所处格子的号码.

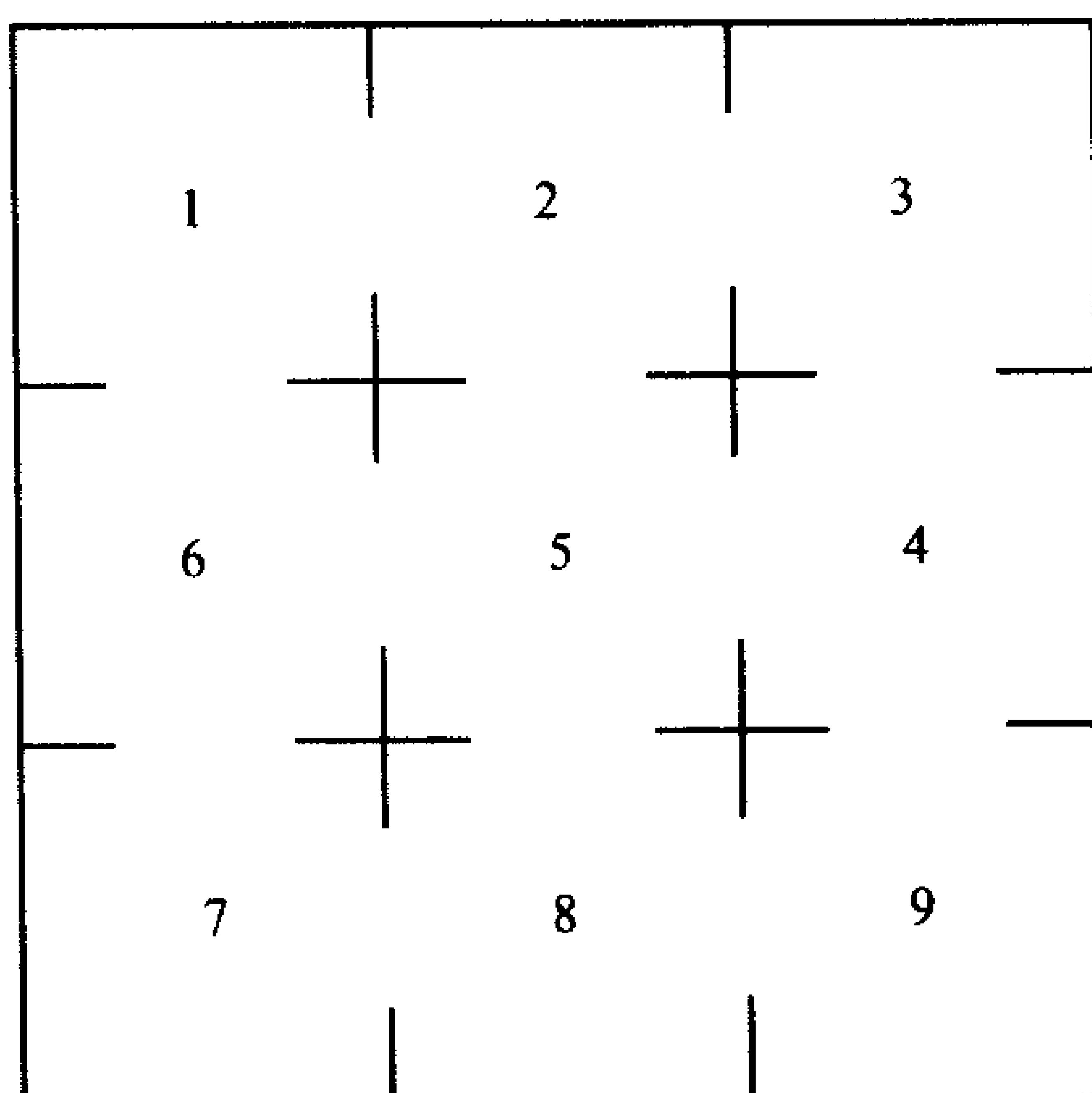
答案:

(a) P_{jk} = Pr{ 在 $n+1$ 次抛掷后正面数减去反面数等于 k| 在 n 次抛掷后正面数减去反面数等于 j}

$$= \begin{cases} p & \text{如果} \quad k=j+1, \\ 1-p & \text{如果} \quad k=j-1, \\ 0 & \text{其他}. \end{cases}$$

73

(b) P_{jk} = Pr{ 在第 $n+1$ 次交换后第一个罐里有 k 个白球 | 在第 n 次交换后第一个罐里有 j个白球}



$$
= \begin{cases} \left(\dfrac{j}{N}\right)^2, & \text{如果} k = j-1, \quad j = 1, 2, \cdots, N, \\ 2\left(\dfrac{j}{N}\right)\left(\dfrac{N-j}{N}\right), & \text{如果} k = j, \quad j = 0, 1, \cdots, N, \\ \left(1 - \dfrac{j}{N}\right)^2, & \text{如果} k = j+1, \quad j = 0, 1, \cdots, N-1, \\ 0 & \text{其他}, \end{cases}
$$

并且 P_{jk} 独立于 n.

2. (a) 罐 A 和 B 共装有 N 个球. 我们做这样的试验：在时刻 $t(t = 1, 2, \cdots)$ 从 N 个球中随机抽取一个球 (所有选取是等可能的), 然后随机抽取一个罐 (选中罐 A 的概率为 p, 选中罐 B 的概率为 q), 再把前面抽取的球放入这个罐中. 每次试验后系统的状态由罐 A 中球的数目来表示. 试确定此马尔可夫链的转移概率矩阵.

(b) 假设在时刻 t 罐 A 中刚好有 k 个球. 在时刻 $t+1$, 一个罐被选取的概率正比于它所装的球数 (即 A 被选中的概率为 k/N, B 被选中的概率为 $(N-k)/N$.) 然后以概率 p 从 A 抽取一个球或以概率 q 从 B 抽取一个球, 放入前面选取的罐中. 试确定此马尔可夫链的转移概率矩阵.

(c) 假设在时刻 $t+1$, 一个球和一个罐被选取的概率都取决于罐中球的数目 (即, 从 A 中抽取球的概率为 $\dfrac{k}{N}$, 从 B 中抽取球的概率为 $\dfrac{N-k}{N}$, 罐 A 以概率 $\dfrac{k}{N}$ 被选中, 罐 B 以概率 $\dfrac{N-k}{N}$ 被选中). 该系统的状态为罐 A 所含球的数量, 试确定该马尔可夫链的转移概率矩阵.

(d) 试确定 (a)、(b) 和 (c) 中的等价类.

答案：

$$
\text{(a) } P_{ik}\begin{cases}\dfrac{N-i}{N}p, & \text{若}\quad k=i+1,\\ \dfrac{i}{N}p+\dfrac{N-i}{N}q, & \text{若}\quad k=i,\\ \dfrac{i}{N}q, & \text{若}\quad k=i-1,\\ 0, & \text{其他},\end{cases}\qquad i=0,1,2,\cdots,N.
$$

74

一个等价类是 $\{0,1,2,\cdots,N\}$.

$$
\text{(b) } P_{ik}\begin{cases}\dfrac{i}{N}q, & \text{若}\quad k=i+1,\\ \dfrac{i}{N}p+\dfrac{N-i}{N}q, & \text{若}\quad k=i,\\ \dfrac{N-i}{N}p, & \text{若}\quad k=i-1,\\ 0, & \text{其他},\end{cases}\qquad i=1,2,\cdots,N-1.
$$

$P_{ii}=1$ 若 $i=0$ 和 $i=N$. 等价类是 $\{0\},\{N\},\{1,2,\cdots,N-1\}$.

$$
\text{(c) } P_{ik}\begin{cases}\dfrac{i^2}{N^2}+\dfrac{(N-i)^2}{N^2}, & \text{若}\quad k=i,\\ \dfrac{i(N-i)}{N^2}, & \text{若}\quad k=i+1,\text{或}k=i-1,\\ 0, & \text{其他},\end{cases}
$$

等价类是 $\{0\},\{1,2,\cdots,N-1\},\{N\}$.

3. (a) 设一个心理被试对一组刺激 $\{S_1,S_2,\cdots,S_N\}$ 的每一个刺激作出 A_1 和 A_2 两个反应中的一个. 现随机抽选单个刺激 (所有刺激被选是等可能的), 被试对所抽选的刺激作出反应. 在每次试验中以概率 $\pi(0<\pi<1)$ 强化, 并且强化与先前试验情况无关. 当强化出现时所选取的刺激不改变它的条件状态, 而当强化不出现时, 刺激却引起另一反应. 考虑该马尔可夫链, 其状态变量是引起反应 A_1 的条件刺激数. 试确定它的转移概率矩阵.

(b) 一心理学被试 S 能作三个反应: A_0,A_1 和 A_2. 反应 A_0 对应于猜测状态. 若 S 作反应 A_1, 则试验以概率 π_1 强化被试, 从而在下一次试验中 S 将作相同的反应. 若强化不出现 (其概率为 $1-\pi_1$) 则下次试验 S 将作猜测反应. 类似地, π_2 是关于反应 A_2 的强化概率. 若强化出现则下一次作相同的反应, 否则将作猜测反应. 当 S 处于猜测状态时, 在下一次试验中仍然处于此状态的概率为 $1-c$, 而分别以 $\dfrac{c}{2}$ 和 $\dfrac{c}{2}$ 的概率作反应 A_1 和 A_2. 考虑由被猜试的状态所构成的马尔可夫链, 试确定其转移概率矩阵.

答案: (a) $P_{ii}=\pi, P_{i,i+1}=\dfrac{N-i}{N}(1-\pi), P_{i,i-1}=\dfrac{i}{N}(1-\pi)$; 其余, $P_{ij}=0(i,j=1,2,\cdots,N)$.

75

(b) $P_{00}=1-c, P_{01}=P_{02}=\dfrac{c}{2}; P_{10}=1-\pi_1, P_{11}=\pi_1, P_{20}=1-\pi_2, P_{22}=\pi_2$.

4. 若马尔可夫链的转移概率矩阵为

(a) $\begin{Vmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{Vmatrix}$, (b) $\begin{Vmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \end{Vmatrix}$.

试确定其不同状态的类和周期.

5. 考虑重复的独立试验, 每次试验分别以概率 p 和 q 出现结果 S(成功) 和 F(失败). 试确定使得首次出现事件 SF(即成功之后紧接着是失败) 所做的试验次数的分布. 若对于事件 SSF 和 SFS, 其结果如何?

6. 可采用如下方法估计某有限群体所含个体的数目, 这对于野生动物群体, 诸如鱼类等可能是恰当的. 先随机抽取该群体中的一个个体, 贴上标签然后放回. 再抽取另外一个, 同样贴上标签后放回, 直至抽取到先前已经抽取过的个体时为止, 假设 T 次抽取之后这种情况发生, 我们则停止试验 (此后还可以加上别的标签重新试验). 我们希望根据 T 的值来估计群体所含个体的数量 N.

设 X_n 是最近连续观察到的没有贴上标签的个体数. 则对于 $n=0,1,\cdots,T-1, X_n=0$, 而 $X_T=0$, 所以 T 是首次达到零的时间, $T=\min\{n\geqslant 1: X_n=0\}$.

(a) 对于任何固定 N, 试证明 (X_n) 是成功游程马尔可夫链, 并指出其概率 p_n 和 q_n.

(b) 计算 $\Pr\{T=t|X_0=0\}, t=2,\cdots,N$.(见第 1 章初等问题 2.)

7. 计算机部件的服役寿命是一个用离散单位度量的随机变量 T, 其中 $\Pr\{T=k\}=a_k, k=1,2,\cdots$. 设从一新部件开始, 每一部件损坏时被新的部件替换. 令 X_n 是正在使用的部件在时刻 n 的年龄, 则 (X_n) 是一成功游程马尔可夫链.

(a) 计算概率 p_i 和 q_i.

"计划替换" 策略是指: 在时刻 N 之前部件损坏时换新的, 在时刻 N 不管部件损坏与否都换新的. 这样, 替换时间是 $T^*=\min\{T,N\}$, 其中 $T=\min\{n\geqslant 1: X_n=0\}$.

(b) 计算 $E[T^*]$. (见第 1 章初等问题 1, 2 和 3.)

8. 设某人身患高传染性疾病未被公共卫生官员察觉而进入一居民点. 在每一段时间里, 该疾病或者以概率 p 传染给另外一个人, 或者由于带病者的症状出现而被公共卫生官员发现, 后者发生的概率为 $1-p$. 试计算在居民点中已被感染但未被发现的患者数目的概率分布. 此处我们假设每次传染都与第一次一样.

76

问 题

1. 每个 $n\times n$ 随机矩阵 $\boldsymbol{P}$(矩阵 $\boldsymbol{P}=\|P_{ij}\|$ 若满足 $0\leqslant P_{ij}\leqslant 1$ 和 $\sum\limits_j P_{ij}=1$, 则称为随机矩阵) 对应一个马尔可夫链, 其一步转移矩阵正是 $\boldsymbol{P}$. 然而并不是每一个 $n\times n$ 随机矩阵都是某马尔可夫链的二步转移矩阵. 特殊地试证明一个 2×2 随机矩阵是某马尔可夫链的二步转移矩阵的充分必要条件是它的主对角线各项之和大于或等于 1.

2. 设 $n_1, n_2, \cdots, n_k$ 是具有最大公因子 d 的正整数列. 证明存在一个正整数 M, 使得当 $m \geqslant M$ 时存在一非负整数列 $\{c_j\}_{j=1}^k$ 满足

$$md = \sum_{j=1}^{k} c_j n_j.$$

(下面问题 4 需要这个结论.)

提示: 令 $A = \{n | n = c_1 n_1 + \cdots + c_k n_k, \{c_i\}为非负整数\}$.

令 $B = \{b_1 n_1 + \cdots + b_j n_j | n_1, n_2, \cdots, n_j \in A, 且 b_1, \cdots, b_j 是正的或负的整数\}$.

设 d' 是 B 中最小的正整数, 证明 d' 是 A 中所有整数的公因子. 然后证明 d' 是 A 中所有整数的最大公因子. 因此 $d' = d$. 重新排列表达式 $d = a_1 n_1 + \cdots + a_l n_l$ 中的各项, 使得正系数项都在前面. 这样 $d = N_1 - N_2$, 且 $N_1 \in A, N_2 \in A$. 令 M 是正整数, $M = N_2^2/d$, 每个整数 $m \geqslant M$ 可表示为 $m = M + k = N_2^2/d + k, (k = 0, 1, 2, \cdots)$, 且 $k = \delta N_2/d + b$, 其中 $0 \leqslant b < N_2/d$ 且 $\delta = j$ 当 $j(N_2/d) \leqslant k < (j+1)N_2/d, j = 0, 1, 2, \cdots$, 所以 $md = N_2^2 + (\delta N_2/d + b)d = N_2(N_2 + \delta - b) + bN_1$.

3. 证明定理 4.1.

提示: 设 $P_{ii}^s > 0, P_{jj}^{n+s+m} \geqslant P_{ji}^n P_{ii}^s P_{ij}^m > 0$ 对某 $m > 0$ 和 $n > 0$. 由于还有 $P_{ii}^{2s} > 0$, 我们有 $P_{jj}^{n+2s+m} > 0$. 这样 $d(j)$ 除尽 $(n+2s+m) - (n+s+m) = s$.

4. 证明定理 4.2 和推论 4.1.

提示: 由问题 2 知存在 N 使得当 $n \geqslant N$ 时, 有

$$P_{ii}^{nd(i)} = P_{ii}^{(c_1 n_1 + \cdots + c_k n_k)}.$$

5. 任给一个有限非周期不可约马尔可夫链, 证明对于某个 n, $\boldsymbol{P}^n$ 所有各项都是正的.

6. 若 j 是瞬态, 试证明对所有 i 有

$$\sum_{n=1}^{\infty} P_{ij}^n < \infty.$$

77

提示: 利用关系式 (5.10).

7. 设一马尔可夫链包含 r 个状态. 试证明:

(a) 若状态 k 能由 j 达到, 则它至多只要 $r-1$ 步便达到;

(b) 若 j 是常返状态, 则存在 $\alpha(0 < \alpha < 1)$ 使得对于任意 $n > r$, 在 n 步转移之后首次返回状态 j 的概率小于等于 α^n.

8. 考虑某贝努利试验序列 $X_1, X_2, X_3, \cdots$, 其中 $X_n = 1$ 或 0. 假设

$$\Pr\{X_n = 1 | X_1, X_2, \cdots, X_{n-1}\} \geqslant \alpha > 0, n = 1, 2, \cdots.$$

证明, (a) $\Pr\{X_n = 1, 对某n\} = 1$, (b) $\Pr\{X_n = 1, 无限经常\} = 1$.

9. 设

$$\boldsymbol{P}=\begin{Vmatrix}1-a & a\\ b & 1-b\end{Vmatrix},\quad 0<a<1,0<b<1.$$

证明

$$\boldsymbol{P}^n=\frac{1}{a+b}\begin{Vmatrix}b & a\\ b & a\end{Vmatrix}+\frac{(1-a-b)^n}{a+b}\begin{Vmatrix}a & -a\\ -b & b\end{Vmatrix}.$$

10. 考虑整数集上的某一随机游动, 对于所有 $i, P_{i,i+1}=p, P_{i,i-1}=q(0<p<1, p+q=1)$. 试确定 P_{00}^n.

答案: $P_{00}^{2m}=\begin{pmatrix}2m\\ m\end{pmatrix}p^m q^m, P_{00}^{2m+1}=0.$

11. (续上) 求 $u_n=P_{00}^n$ 的母函数, 即确定

$$P(x)=\sum_{n=0}^{\infty}u_n x^n.$$

提示: 利用恒等式 $\begin{pmatrix}2n\\ n\end{pmatrix}=(-1)^n\begin{pmatrix}-\frac{1}{2}\\ n\end{pmatrix}2^{2n}$, 其中对任意实数 a 我们定义

$$\begin{pmatrix}a\\ n\end{pmatrix}=\frac{a(a-1)\cdots(a-n+1)}{n!}.$$

答案: $P(x)=(1-4pqx^2)^{-\frac{1}{2}}$.

12. (续上) 试确定从状态 0 出发返回状态 0 时间的母函数.

答案: $F(x)=1-\sqrt{(1-4pqx^2)}$.

13. (续上) 最终返回原点的概率是多少?

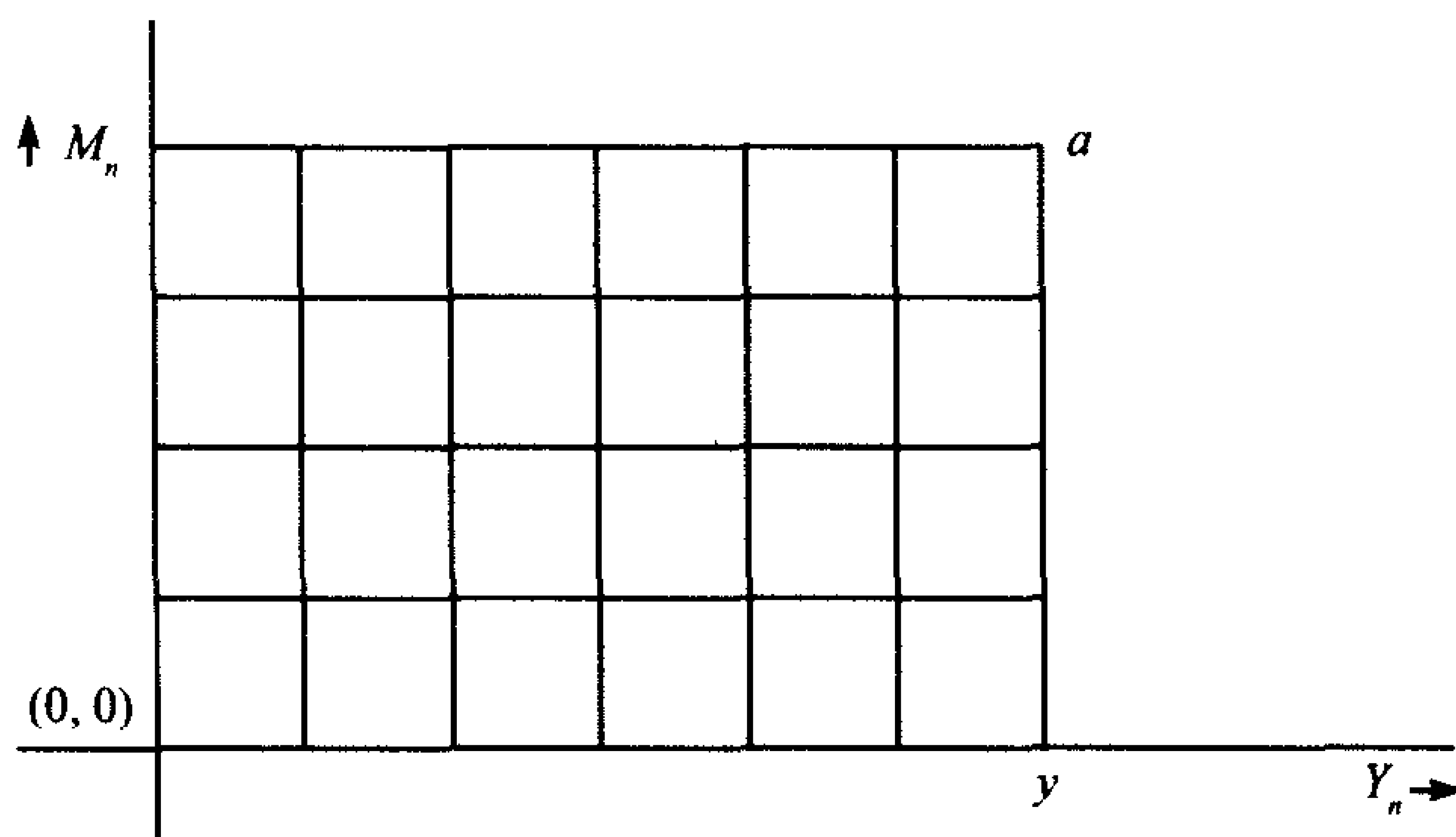
14. 假设两枚不同的均匀硬币被同时反复抛掷, 并把出现正反面的次数记录下来. 考虑在第 n 次抛掷时两枚硬币正面出现的累计次数恰好相等的事件 E_n. 试将事件 E_n 与整数集上对称随机游动某状态的返回时间联系起来.

78

15. 设 $X_1, X_2, \cdots$ 相互独立, 且有 $\Pr\{X_k=+1\}=p, \Pr\{X_k=-1\}=q=1-p$, 其中 $p\geqslant q$. 令 $S_0=0, S_n=X_1+\cdots+X_n$, 则 $M_n=\max\{S_k: 0\leqslant k\leqslant n\}$ 且 $Y_n=M_n-S_n$, 如果 $T(a)=\min\{n: S_n=a\}$, 证明

$$\Pr\left\{\max_{0\leqslant k\leqslant T(a)}Y_k<y\right\}=\begin{cases}\left(\dfrac{y}{1+y}\right)^a, & 若p=q=\dfrac{1}{2},\\[2ex] \left[\dfrac{\dfrac{p}{q}-\left(\dfrac{p}{q}\right)^{y+1}}{1-\left(\dfrac{p}{q}\right)^{y+1}}\right]^a, & 若p\neq q.\end{cases}$$

提示：双变量过程 (M_n, Y_n) 是正整数格上的随机游动. 这个随机游动从项点 a 离开矩形



的概率是多少?

16. (续上) 对上题添加新记号, 令 τ 表示部分和 S_n 与其至时刻 n 为止偏离最大值 y 个单位的首达时间, 即 $\tau = \min\{n : Y_n = y\}$. 试证明 $M_\tau = S_\tau + y$ 具有几何分布 $\Pr\{M_\tau \geqslant a\} = \theta^a$, 其中 $a = 1, 2, \cdots$, 并确定 θ.

提示：$M_\tau \geqslant a$ 当且仅当 $\max\limits_{0 \leqslant k \leqslant T(a)} Y_k < y$.

附 记

马尔可夫链的部分理论可在 Feller[1] 的后半部分找到.

79

Kemeny 和 Snell 著的书 [2] 有许多关于马尔可夫链的生动例子, 它们来自心理学、社会学、经济学、生物学等学科.

Chung[3] 是致力于马尔可夫链结构分析方面的最高等的论著.

参考书目

[1] W.Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 2nd. ed. Wiley, New York, 1957.

[2] J.G. Kemeny and J.L.Snell, *Finite Markov Chains.* Van Nostrand, Princeton, New Jersey, 1960.

[3] K.L. Chung, *Markov Chains with Stationary Transition Probabilities,* Springer-Verlag,

80

Berlin, 1960.

第 3 章　马尔可夫链的基本极限定理和应用

本章 3.1~3.4 节是研究马尔可夫链的重要工具, 每一本入门性的教科书都有这一内容. 然而, 读者可以把 3.2 节关于离散更新定理的证明推延到第 5 章阅读, 在第 5 章将介绍一般的更新理论.

3.5~3.6 节中的例子在随机排队模型中是经典的. 其中一部分内容第一次阅读时可以略过.

3.1　离散更新方程

下面定理提供了马尔可夫链分析的关键工具.

定理 1.1　设三序列 $\{a_k\},\{u_k\},\{b_k\},k=0,\pm1,\pm2,\cdots$, 满足条件 $a_k\geqslant 0$, $\sum a_k=1,\sum|k|a_k<\infty,\sum ka_k>0,\sum|b_k|<\infty$, 并且, 使 $a_k>0$ 的所有整数 k 的最大公约数为 1. 如果有界实数列 $\{u_n\}$ 满足更新方程

$$u_n-\sum_{k=-\infty}^{+\infty}a_{n-k}u_k=b_n,\quad n=0,\pm1,\pm2,\cdots,$$

则 $\lim\limits_{n\to\infty}u_n$ 和 $\lim\limits_{n\to-\infty}u_n$ 存在. 再者, 如果

$$\lim_{n\to-\infty}u_n=0,$$

则

$$\lim_{n\to\infty}u_n=\frac{\sum\limits_{k=-\infty}^{\infty}b_k}{\sum\limits_{k=-\infty}^{\infty}ka_k}.\tag{1.1}$$

在 $\sum\limits_{k=-\infty}^{\infty}ka_k=\infty$ 的情况下, 上述极限关系式仍然成立, 只是这时我们认定

$$\frac{\sum\limits_{k=-\infty}^{\infty}b_k}{\sum\limits_{k=-\infty}^{\infty}ka_k}=0.$$

这个定理的一般形式的证明已超出本书范围. 实际上我们仅在 k 为负值时 $\{a_k\}$, $\{u_k\}, \{b_k\}$ 为 0 且 $b_k \geqslant 0$ 的情况下利用此定理, 此情况下的定理证明将在 3.2 节叙述.

附注 1.1 在 $a_{-k}=0, b_{-k}=0$ 且 $u_{-k}=0(k>0)$ 时, 更新方程变为

$$u_n - \sum_{k=0}^{n} a_{n-k}u_k = b_n, \quad n=0,1,2,\cdots.$$

附注 1.2 (使用"更新方程"一语的由来) 考虑一个灯泡的寿命, 用离散单位来测量, 它是一随机变量 ξ, 其分布为

$$\Pr\{\xi = k\} = a_k, \quad k=0,1,2,\cdots, a_k>0, \quad \sum_{k=0}^{\infty} a_k = 1.$$

每个灯泡用坏后即换一个新灯泡. 设第一个灯泡使用时间持续至 ξ_1, 第二个灯泡持续至 $\xi_1+\xi_2$, 第 n 个灯泡使用时间持续至 $\sum_{i=1}^{n}\xi_i$, 这里 ξ_i 是独立同分布随机变量, 每个分布都与 ξ 相同. 令 u_n 表示至时刻 n 为止更新数目的期望值. 如果第一次更换出现在时刻 k, 那么到时刻 n 为止的余下时间的更新数目的期望值为 u_{n-k}, 对所有可能值 k 求和, 得

$$\begin{aligned} u_n &= \sum_{k=0}^{n}(1+u_{n-k})a_k + 0\sum_{k=n+1}^{\infty} a_k \\ &= \sum_{k=0}^{n} u_{n-k}a_k + \sum_{k=0}^{n} a_k \\ &= \sum_{k=0}^{n} a_{n-k}u_k + b_n, \end{aligned} \tag{1.2}$$

其中

$$\sum_{k=0}^{n} a_k = b_n.$$

式子 (1.2) 成立的理由如下：若第一个灯泡在时刻 $k(0 \leqslant k \leqslant n)$ 坏掉, 时刻 n 之前更新数目的期望值为 $1+u_{n-k}$, 而这个事件概率为 a_k. 第二个和式表示第一个灯泡持续使用时间超过时刻 n 的概率. 注意到过程的再生性质, 我们可把事件按所有可能的第一次更换时间进行分解求得 u_n.

82

下面定理实际上是在这种特殊情况下的遍历性定理, 描述了非周期常返马尔可夫链对所有 i 和 j 当 $n\to\infty$ 时 P_{ij}^n 的极限性质. 它的证明是基本更新极限定理的一个简单应用.

定理 1.2 (马尔可夫链基本极限定理)

(a) 考虑一个常返不可约非周期马尔可夫链. 令 P_{ii}^n 表示在给定 $X(0)=i$(初始状态为 i) 的条件下经过 n 次转移到达状态 i 的概率, $n=0,1,2\cdots$. 按前述规定 $P_{ii}^0=1$. 设 f_{ii}^n 表示在第 n 次转移首返状态 i 的概率, $n=0,1,2,\cdots$, 其中 $f_{ii}^0=0$. 于是

$$P_{ii}^n-\sum_{k=0}^{n}f_{ii}^{n-k}P_{ii}^k=\begin{cases}1, & 若\ n=0,\\ 0, & 若\ n>0.\end{cases}$$

[这就是第 2 章已导出的公式 (5.1).] 则

$$\lim_{n\to\infty}P_{ii}^n=\frac{1}{\sum\limits_{n=0}^{\infty}nf_{ii}^n}.$$

(b) 在 (a) 的相同条件下, $\lim\limits_{n\to\infty}P_{ji}^n=\lim\limits_{n\to\infty}P_{ii}^n$ 成立.

证明 (a) 令

$$u_n=P_{ii}^n,\quad n\geqslant 0;\quad u_n=0,\quad n<0;$$

$$a_n=f_{ii}^n,\quad n\geqslant 0;\quad a_n=0,\quad n<0;$$

$$b_n=\begin{cases}1, & n=0,\\ 0, & n\neq 0;\end{cases}$$

然后应用定理 1.1.

(b) 我们利用递推关系式

$$P_{ji}^n=\sum_{u=0}^{n}f_{ji}^uP_{ii}^{n-u},\quad i\neq j,\quad n\geqslant 0$$

[参阅第 2 章公式 (5.9)]. 更一般地, 令

$$y_n=\sum_{k=0}^{n}a_{n-k}x_k,$$

其中

$$a_m\geqslant 0,\quad \sum_{m=0}^{\infty}a_m=1,\quad \lim_{k\to\infty}x_k=c.$$

在这些情况下, 我们证明 $\lim\limits_{n\to\infty}y_n=c$. 事实上,

$$y_n-c=\sum_{k=0}^{n}a_{n-k}x_k-c\sum_{m=0}^{\infty}a_m=\sum_{k=0}^{n}a_{n-k}(x_k-c)-c\sum_{m=n+1}^{\infty}a_m.$$

83

对于事先指定的 $\varepsilon > 0$, 我们确定 $K(\varepsilon)$ 使得 $|x_k - c| < \dfrac{\varepsilon}{3}$ 对所有 $k \geqslant K(\varepsilon)$ 成立.

$$y_n - c = \sum_{k=0}^{K(\varepsilon)} a_{n-k}(x_k - c) + \sum_{k=K(\varepsilon)+1}^{n} a_{n-k}(x_k - c) - c\sum_{m=n+1}^{\infty} a_m,$$

所以

$$|y_n - c| \leqslant M\sum_{k=0}^{K(\varepsilon)} a_{n-k} + \frac{\varepsilon}{3}\sum_{h=K(\varepsilon)+1}^{n} a_{n-h} + |c|\sum_{m=n+1}^{\infty} a_m,$$

其中

$$M = \max_{k\geqslant 0} |x_k - c|.$$

我们选择 $N(\varepsilon)$ 使得 $|c|\displaystyle\sum_{m=n+1}^{\infty} a_m < \frac{\varepsilon}{3}$ 且对于 $n \geqslant N(\varepsilon)$,

$$\sum_{k=0}^{k(\varepsilon)} a_{n-k} \equiv \sum_{m=n-K(\varepsilon)}^{n} a_m < \frac{\varepsilon}{3M}.$$

则, 当 $n \geqslant N(\varepsilon)$ 时,

$$|y_n - c| \leqslant \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

现在令

$$y_n = P_{ji}^n, \quad a_n = f_{ji}^n, \quad x_n = P_{ii}^n,$$

则得到想要的结果.

附注 1.3 设 C 是一个常返类. 则对于 $i \in C, j \notin C$, 及每个 $n, P_{ij}^n = 0$. 因此状态一旦进入 C 就不可能再离开 C. 由此推知, 子矩阵 $\|P_{ij}\|, i, j \in C$ 是转移概率矩阵, 它所联系的马尔可夫链是不可约的且常返的. 故极限定理同样适用任何非周期常返类.

附注 1.4 如果当 $n \to \infty$ 时 $a_n \to a$, 则由初等方法可以证明

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} a_k = a. \tag{1.3}$$

这样, 如果 i 属于某个常返非周期类, 则

$$\lim_{n\to\infty} \frac{1}{n}\sum_{m=1}^{n} P_{ii}^m = \frac{1}{\displaystyle\sum_{n=0}^{\infty} n f_{ii}^n} = \frac{1}{m_i}, \tag{1.4}$$

84

其中 m_i 是平均返回时间.

如果 i 属于某个常返周期类且周期为 d, 则可以证明 (见第 2 章问题 4) 当 m 不是 d 的倍数 (即对于任意整数 $n, m \neq nd$) 时, $P_{ii}^m = 0$, 并且

$$\lim_{n\to\infty} P_{ii}^{nd} = \frac{d}{m_i}.$$

结合式 (1.3) 后这两个结果, 表明式 (1.4) 对周期情况也成立.

如果对于某个属于非周期常返类的状态 i, $\lim_{n\to\infty} P_{ii}^n = \pi_i > 0$, 则对于 i 所属类中的所有 $j, \pi_j > 0$ 成立 (这个事实的证明可仿效第 2 章推论 5.1 的方法, 此处省略). 在这种情况下, 我们称该类为**正常返的或强遍历的**. 如果每个 $\pi_i = 0$, 且类是常返的, 我们称其为**零常返的或弱遍历的**.

定理 1.3 在一个具有状态 $j = 0, 1, 2, \cdots$ 的正常返非周期类中,

$$\lim_{n\to\infty} P_{jj}^n = \pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}, \quad \sum_{i=0}^{\infty} \pi_i = 1,$$

成立, 且 π_i 由下列方程组唯一确定

$$\pi_i \geqslant 0, \quad \sum_{i=0}^{\infty} \pi_i = 1, \quad \pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}. \tag{1.5}$$

满足 (1.5) 的任意集合 $(\pi_i)_{i=0}^{\infty}$ 称为马尔可夫链的**平稳概率分布**. 我们在第 11 章将进一步研究这个概念.

证明 对每个 n 和 M, $1 = \sum_{j=0}^{\infty} P_{ij}^n \geqslant \sum_{j=0}^{M} P_{ij}^n$. 令 $n \to \infty$ 并利用定理 1.2, 对任何 M 我们得到 $1 \geqslant \sum_{j=0}^{M} \pi_j$. 于是 $\sum_{j=0}^{\infty} \pi_j \leqslant 1$. 由于 $P_{ij}^{n+1} \geqslant \sum_{k=0}^{M} P_{ik}^n P_{kj}$, 如果令 $n \to \infty$, 则我们得 $\pi_j \geqslant \sum_{k=0}^{M} \pi_k P_{kj}$. 接下来, 由于左边与 M 无关, 当 $M \to \infty$ 时, 有

$$\pi_j \geqslant \sum_{k=0}^{\infty} \pi_k P_{kj}. \tag{1.6}$$

用 P_{ji} 乘上式两端, 然后依 j 求和并利用式 (1.6), 推得 $\pi_j \geqslant \sum_{k=0}^{\infty} \pi_k P_{kj}^2$. 如此类推, 对任意 n 有 $\pi_j \geqslant \sum_{k=0}^{\infty} \pi_k P_{kj}^n$. 假设严格不等式对于某个 j 成立, 关于 j 把这些不等式相加, 我们有

$$\sum_{j=0}^{\infty} \pi_j > \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \pi_k P_{kj}^n = \sum_{k=0}^{\infty} \pi_k \sum_{j=0}^{\infty} P_{kj}^n = \sum_{k=0}^{\infty} \pi_k,$$

85

这就发生了矛盾. 因此对所有 $n, \pi_j = \sum_{k=0}^{\infty} \pi_k P_{kj}^n$. 令 $n \to \infty$, 由于 $\sum \pi_k$ 收敛且 P_{kj}^n 是一致有界的, 则对每个 j 有

$$\pi_j = \sum_{k=0}^{\infty} \pi_k \lim_{n\to\infty} P_{kj}^n = \pi_j \sum_{k=0}^{\infty} \pi_k.$$

由正常返性知 $\pi_j > 0$, 因此 $\sum_{k=0}^{\infty} \pi_k = 1$.

假设 $x = \{x_n\}$ 满足关系式 (1.5), 则

$$x_k = \sum_{j=0}^{\infty} x_j P_{jk} = \sum_{j=0}^{\infty} x_j P_{jk}^n,$$

如令 $n \to \infty$, 得

$$x_k = \sum_{j=0}^{\infty} x_j \lim_{n\to\infty} P_{jk}^n = \pi_k \sum_{j=0}^{\infty} x_j = \pi_k.$$

■

例 考虑一类随机游动, 其转移矩阵为 (参阅 2.2 节例 B)

$$\boldsymbol{P} = \|P_{ij}\| = \left\| \begin{matrix} 0 & 1 & 0 & \cdots & \cdots \\ q_1 & 0 & p_1 & \cdots & \cdots \\ 0 & q_2 & 0 & p_2 & \cdots \\ \vdots & & & & \end{matrix} \right\|.$$

这个马尔可夫链具有周期 2. 然而, 这一次我们要研究其平稳概率分布的存在性, 也就是说, 我们希望确定方程

$$x_i = \sum_{j=0}^{\infty} x_j P_{ji} = p_{i-1} x_{i-1} + q_{i+1} x_{i+1}, i = 0, 1, \cdots, \tag{1.7}$$

满足规范化条件 $\sum_{i=0}^{\infty} x_i = 1$ 的正解, 其中 $p_{-1} = 0, p_0 = 1$, 因此 $x_0 = q_1 x_1$. 在方程 (1.7) 中令 $i = 1$, 我们可以用 x_0 确定 x_2. 再令 $i = 2$, 可以用 x_0 确定 x_3. 如此类推, 得

$$x_i = \frac{p_{i-1} p_{i-2} \cdots p_1}{q_i q_{i-1} \cdots q_1} x_0 = x_0 \prod_{k=0}^{i-1} \frac{p_k}{q_{k+1}}. \quad i \geqslant 1,$$

是方程 (1.7) 的解, 其中 x_0 待确定. 由于

$$1 = x_0 + \sum_{i=1}^{\infty} x_0 \prod_{k=0}^{i-1} \frac{p_k}{q_{k+1}},$$

我们有

$$x_0 = \frac{1}{1+\sum_{i=1}^{\infty}\prod_{k=0}^{i-1}\frac{p_k}{q_{k+1}}},$$

所以

$$x_0 > 0\text{当且仅当 } \sum_{i=0}^{\infty}\prod_{k=0}^{i-1}\frac{p_k}{q_{k+1}} < \infty.$$

特别地, 如果对于 $k \geqslant 1, p_k = p$ 且 $q_k = q = 1-p$, 则仅当 $p < q$ 时, 级数

$$\sum_{i=1}^{\infty}\prod_{k=0}^{i-1}\frac{p_k}{q_{k+1}} = \frac{1}{p}\sum_{i=1}^{\infty}\left(\frac{p}{q}\right)^i$$

收敛.

3.2 定理 1.1 的证明

现在我们对 $k<0$ 时 a_k, b_k, u_k 均为 0, $k \geqslant 0$ 时 $a_k, b_k \geqslant 0$ 及 $a_1 > 0, \sum_{k=0}^{\infty} a_k = 1$ 的情况证明定理 1.1. 这时更新方程变为

$$u_n - \sum_{k=0}^{n} a_{n-k}u_k = b_n, \quad n = 0,1,2,\cdots,$$

或等价地

$$u_n - \sum_{k=0}^{n} a_k u_{n-k} = b_n, \quad n = 0,1,2,\cdots. \tag{2.1}$$

逐个地考虑方程, 容易归纳推得对所有 $k, u_k \geqslant 0$.

由于假设 $\{u_n\}$ 是有界序列, 故 $\lambda = \limsup_{n\to\infty} u_n$ 是有限的. 令 $n_1 < n_2 < \cdots$ 表示一个子序列满足 $\lim_{j\to\infty} u_{n_j} = \lambda$. 我们应用条件 $a_1 > 0$ 证明 $\lim_{j\to\infty} u_{n_j-1} = \lambda$. 用反证法, 假设最后的关系式不成立, 则由 λ 的定义可知存在一个 $\lambda' < \lambda$ 使得对无限多个 $j, u_{n_j-1} < \lambda'$. 取 $\varepsilon = [a_1(\lambda-\lambda')]/4, M = \sup_{n\geqslant 0} u_n$, 并确定 N 使得当 $n \geqslant N$ 时

$$\sum_{k=0}^{n} a_k > 1 - \frac{\varepsilon}{M}. \tag{2.2}$$

选择充分大的 j 使得 $n_j \geqslant N$ 且

$$u_{nj} > \lambda - \varepsilon, \quad u_{n_j-1} < \lambda' < \lambda, \quad 0 \leqslant b_{n_j} < \varepsilon$$

及

87

$$\text{对所有} n \geqslant n_j - N, \quad u_n < \lambda + \varepsilon. \tag{2.3}$$

由 λ 的定义和 λ' 的意义可知, 上述选取均可做到. 由式 (2.1), 式 (2.2) 和式 (2.3), 我们有

$$\begin{aligned}
u_{n_j} &\leqslant \sum_{k=0}^{n_j} a_k u_{n_j-k} + \varepsilon < \sum_{k=0}^{N} a_k u_{n_j-k} + M \sum_{k=N+1}^{n_j} a_k + \varepsilon \\
&< \sum_{k=0}^{N} a_k u_{n_j-k} + 2\varepsilon && [\text{利用式 (2.2) 和式 (2.3)}] \\
&< (a_0 + a_2 + a_3 + \cdots + a_{N-1} + a_N)(\lambda + \varepsilon) + a_1\lambda' + 2\varepsilon && [\text{利用式 (2.3)}] \\
&\leqslant (1 - a_1)(\lambda + \varepsilon) + a_1\lambda' + 2\varepsilon < \lambda + 3\varepsilon - a_1(\lambda - \lambda') = \lambda - \varepsilon.
\end{aligned}$$

最后一行用到了 ε 的选取. 但这和式 (2.3) 中第一个不等式矛盾, 所以 $\lim_{j\to\infty} u_{n_j-1} = \lambda$.

重复上述论证, 我们推得, 对任何非负整数 d 都有

$$\lim_{j\to\infty} u_{n_j-d} = \lambda. \tag{2.4}$$

其次令 $r_n = a_{n+1} + a_{n+2} + \cdots$, 显然, $\sum_{k=0}^{\infty} k a_k = \sum_{n=0}^{\infty} r_n$, 这只需要验证其部分和相等即可. (我们不必假定级数 $\sum r_n$ 收敛.) 此外, $a_1 = r_0 - r_1, a_2 = r_1 - r_2$, 等等, 把它们代入式 (2.1), 我们有

$$\begin{aligned}
&r_0 u_n + r_1 u_{n-1} + \cdots + r_n u_0 \\
&= r_0 u_{n-1} + r_1 u_{n-2} + \cdots + r_{n-1} u_0 + b_n, \quad n = 1, 2, \cdots.
\end{aligned}$$

记 $A_n = r_0 u_n + \cdots + r_n u_0$, 我们把上式写成

$$A_n = A_{n-1} + b_n, \quad n = 1, 2, \cdots,$$

其中 $A_0 = r_0 u_0 = (1 - a_0) u_0 = b_0$. 由此推得 $A_n = \sum_{i=0}^{n} b_i$. 由于 $r_n \geqslant 0$ 和 $u_n \geqslant 0$ 对所有 n 成立, 对任何固定的 $N > 0$ 和 $j > 0$ 我们得到

$$r_0 u_{n_j} + r_1 u_{n_j-1} + \cdots + r_N u_{n_j-N} \leqslant A_{n_j} = \sum_{n=0}^{n_j} b_n.$$

令 $j \to \infty$ 则得关系式 $(r_0 + \cdots + r_N)\lambda \leqslant \sum_{n=0}^{\infty} b_n$, 或等价地, $\lambda \leqslant \sum_{n=0}^{\infty} b_n \left(\sum_{n=0}^{N} r_n\right)^{-1}$.

由 $N>0$ 的任意性, 推得

$$\lambda \leqslant \frac{\sum_{n=0}^{\infty} b_n}{\sum_{n=0}^{\infty} r_n}. \tag{2.5}$$

由于 $u_k \geqslant 0$ 对所有的 k 成立, 这就证明了定理在情况 $\sum_{n=0}^{\infty} r_n = \infty$ 下成立, 因为由式 (2.5) 易得 $\lambda = \lim_{n\to\infty} u_n = 0$.

88

若 $\sum_{n=0}^{\infty} r_n < \infty$. 令 $\mu = \liminf_{n\to\infty} u_n$. 仿照 lim sup 情况的推理我们得知, 如果 $\lim_{j\to\infty} u_{n_j} = \mu$, 则对每个非负整数 d, $\lim_{j\to\infty} u_{n_j-d} = \mu$. 记 $\sum_{n=N+1}^{\infty} r_n = g(N)$. 那么显然有 $\lim_{N\to\infty} g(N) = 0$ 且

$$\sum_{n=0}^{n_j} b_n \leqslant r_0 u_{n_j} + r_1 u_{n_j-1} + \cdots + r_N u_{n_j-N} + g(N)\cdot M.$$

令 $j\to\infty$, 我们推得 $\sum_{n=0}^{\infty} b_n \leqslant (r_0+\cdots+r_N)\mu + g(N)\cdot M$.

现在, 当 $N\to\infty$ 时取极限, 得到

$$\sum_{n=0}^{\infty} b_n \leqslant \mu \sum_{n=0}^{\infty} r_n \quad \text{或} \quad \mu \geqslant \frac{\sum_{n=0}^{\infty} b_n}{\sum_{n=0}^{\infty} r_n}. \tag{2.6}$$

但由式 (2.5) 和式 (2.6) 结合则有 $\mu \geqslant \lambda$. 另一方面, 由定义有 $\mu \leqslant \lambda$. 于是 $\mu = \lambda$, 这就意味着 $\lim_{n\to\infty} u_n$ 存在且它的值是

$$\lim_{n\to\infty} u_n = \frac{\sum_{n=0}^{\infty} b_n}{\sum_{n=0}^{\infty} r_n}.$$

对于 $a_1 = 0$ 且使 $a_m > 0$ 的所有 m 的最大公约数为 1 的情况, 可借助第 2 章推论 4.1 和上述方法完成定理的证明.

3.3 吸收概率

前面我们已经证明了 (第 2 章问题 6): 若 j 是瞬态, 则 $P_{ij}^n \to 0$; 若 i, j 是属于

同一非周期常返类, 则 $P_{ij}^n \to \pi_j \geqslant 0$, 如果 i, j 属于同一周期常返类, 则用 $n^{-1}\sum_{m=1}^{n} P_{ij}^m$ 代替上述的 P_{ij}^n, 相同的结论也成立. 为了完成 P_{ij}^n 极限性质的讨论, 下面考虑 i 是瞬态而 j 为常返状态的情况.

如果 T 是所有瞬态的集合, 我们考虑

$$x_i^1 = \sum_{j \in T} P_{ij} \leqslant 1, i \in T,$$

并归纳定义

$$x_i^n = \sum_{j \in T} P_{ij} x_j^{n-1}, \quad n \geqslant 2, i \in T.$$

注意, x_i^n 正好是从状态 i 出发经过 n 步转移其状态仍然属于 T 的概率. 由于 $x_i^n \leqslant 1$ 对所有 $n \geqslant 1$ 成立 (因为它们都是概率), 我们通过归纳法可以证明 x_i^n 作为n 的函数是单调不增的. 事实上,

89

$$x_i^2 = \sum_{j \in T} P_{ij} x_j^1 \leqslant \sum_{j \in T} P_{ij} = x_i^1.$$

现在我们假设对所有 $j \in T, x_j^n \leqslant x_j^{n-1}$, 我们有

$$0 \leqslant x_i^{n+1} = \sum_{j \in T} P_{ij} x_j^n \leqslant \sum_{j \in T} P_{ij} x_j^{n-1} = x_i^n.$$

因此, $x_i^n \downarrow x_i$, 即 x_i^n 递减趋向于某个极限 x_i, 并且,

$$x_i = \sum_{j \in T} P_{ij} x_j, \qquad i \in T. \tag{3.1}$$

由此推得, 如果上述方程组唯一的有界解是零向量 $(0, 0, \cdots)$, 则从任何瞬态出发被吸收到某常返类将以概率 1 发生. 事实上, 显而易见 $x_i (i \in T)$ 是从状态 i 出发一直未被常返类吸收的概率. 因为这个序列是式 (3.1) 的有界解, 由此可知对所有 $i, x_i = 0$.

附注 3.1　如果一个马尔可夫链仅有有限个 (设为 M) 状态, 则不存在零常返状态, 并且不是所有状态都是瞬态. 事实上, 既然 $\sum_{j=0}^{M-1} P_{ij}^n = 1$ 对所有 n 成立, 因此不可能对所有 j, $\lim_{n\to\infty} P_{ij}^n = 0$ 都成立.

限制于常返类, 相同的论证也说明不存在零状态. 令 $C, C_1, C_2, \cdots$ 表示常返类. 我们定义 $\pi_i(C)$ 为从瞬态 i 出发而最后被常返类 C 吸收的概率. (注意, 一旦过程进入某一常返类, 它就不会再离开这一类了.)

设 $\pi_i^n(C)$ 表示从初始状态 $i(i \in T)$ 出发经过 n 步转移之后第一次进入 C 从而被 C 吸收的概率, 那么

$$\pi_i(C) = \sum_{n=1}^{\infty} \pi_i^n(C) \leqslant 1, \tag{3.2}$$

$$\begin{aligned} \pi_i^1(C) &= \sum_{j \in C} P_{ij}, \\ \pi_i^n(C) &= \sum_{j \in T} P_{ij} \pi_j^{n-1}(C), n \geqslant 2. \end{aligned} \tag{3.3}$$

利用式 (3.3), 式 (3.2) 重写为

$$\begin{aligned} \pi_i(C) &= \pi_i^1(C) + \sum_{n=2}^{\infty} \pi_i^n(C) \\ &= \pi_i^1(C) + \sum_{n=2}^{\infty} \sum_{j \in T} P_{ij} \pi_j^{n-1}(C) \\ &= \pi_i^1(C) + \sum_{j \in T} P_{ij} \sum_{n=2}^{\infty} \pi_j^{n-1}(C), \end{aligned}$$

$$\pi_i(C) = \pi_i^1(C) + \sum_{j \in T} P_{ij} \pi_j(C), i \in T. \tag{3.4}$$

90

假设齐次方程组

$$w_i = \sum_{j \in T} P_{ij} w_j, \quad i \in T$$

唯一的有界解是零向量, 则 $\{\pi_i(C)\}$ 是方程组 (3.4) 的唯一有界解. 此外, 或者对某些 $i \in T, \pi_i^1(C) > 0$, 或者对每个 $i \in T, \pi_i(C) = 0$, 因此对于所有的 n 都有 $\pi_i^n(C) = 0$.

定理 3.1 设 $j \in C$(C 非周期常返类). 则对于 $i \in T$, 我们有

$$\lim_{n \to \infty} P_{ij}^n = \pi_i(C) \lim_{n \to \infty} P_{jj}^n = \pi_i(C) \pi_j.$$

证明 显然 $\pi_i^n(C) = \sum\limits_{k \in C} \pi_{ik}^n(C)$, 其中 $\pi_{ik}^n(C)$ 表示从状态 i 出发经过 n 步转移被吸收到类 C 中并且处于状态 k 的概率. 我们有

$$\pi_i(C) = \sum_{u=1}^{\infty} \sum_{k \in C} \pi_{ik}^u(C) \leqslant 1.$$

所以对任意 $\varepsilon > 0$, 存在有限多个状态的类 $C' \subset C$ 及整数 $N(\varepsilon) = N$, 使得当 $n > N$ 时, 成立

$$\left| \pi_i(C) - \sum_{u=1}^{n} \sum_{k \in C'} \pi_{ik}^u(C) \right| < \varepsilon,$$

即

$$\left|\sum_{u=1}^{\infty}\sum_{k\in C}\pi_{ik}^{u}-\sum_{u=1}^{n}\sum_{k\in C'}\pi_{ik}^{u}\right|<\varepsilon. \tag{3.5}$$

(这里, $\pi_{ik}^{u}(C)$ 简记为 π_{ik}^{u}.)

对于 $j\in C$, 考虑

$$P_{ij}^{n}-\sum_{u=1}^{n}\sum_{k\in C'}\pi_{ik}^{u}\pi_{j}.$$

把事件按照第一次进入 C 中的某状态的时间进行分解, 然后使用通常的递推方法, 我们得到

$$P_{ij}^{n}=\sum_{u=1}^{n}\sum_{k\in C}\pi_{ik}^{u}P_{kj}^{n-u}, i\in T, j\in C.$$

综上所述, 我们有

$$\begin{aligned}
&\left|P_{ij}^{n}-\left(\sum_{u=1}^{n}\sum_{k\in C'}\pi_{ik}^{u}\right)\pi_{j}\right|\\
=&\left|\sum_{u=1}^{n}\sum_{k\in C'}\pi_{ik}^{u}\left(P_{kj}^{n-u}-\pi_{j}\right)+\sum_{u=1}^{n}\sum_{k\in C,k\notin C'}\pi_{ik}^{u}P_{kj}^{n-u}\right|\\
\leqslant&\left|\sum_{u=1}^{N}\sum_{k\in C'}\pi_{ik}^{u}\left(P_{kj}^{n-u}-\pi_{j}\right)\right|\\
&+\left|\sum_{u=N+1}^{n}\sum_{k\in C'}\pi_{ik}^{u}\left(P_{kj}^{n-u}-\pi_{j}\right)\right|+\sum_{u=1}^{n}\sum_{k\in C,k\notin C'}\pi_{ik}^{u}P_{kj}^{n-u}.
\end{aligned}$$

91

但如果 C 是非周期的且 $k\in C'$, 则 $P_{kj}^{n-u}\leqslant 1, \left|P_{kj}^{n-u}-\pi_{j}\right|\leqslant 2$ 且 $\lim\limits_{n\to\infty}P_{kj}^{n-u}=\pi_{j}$. 因此存在 $N'>N$, 使得当 $n>N'$ 时, $\left|P_{kj}^{n-N}-\pi_{j}\right|<\varepsilon(k\in C')$, 所以当 $n>N'$ 时

$$\left|P_{ij}^{n}-\left(\sum_{u=1}^{n}\sum_{k\in C'}\pi_{ik}^{u}\right)\pi_{j}\right|\leqslant\varepsilon+2\sum_{u=N+1}^{n}\sum_{k\in C'}\pi_{ik}^{u}+\sum_{u=1}^{n}\sum_{k\in C,k\notin C'}\pi_{ik}^{u}.$$

然而, 由 N 和 C' 的选择使我们确信右端 $\leqslant 4\varepsilon$. 借助于式 (3.5) 和上述结果, 我们得到

$$\left|P_{ij}^{n}-\pi_{i}(C)\pi_{j}\right|\leqslant 4\varepsilon+\varepsilon\pi_{j},\quad n>N',$$

所以

$$\lim_{n\to\infty}P_{ij}^{n}=\pi_{i}(C)\pi_{j}.$$

■

如果 C 是周期的并且 $j \in C$, 类似方法可用于证明

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P_{ij}^{m} = \pi_i(C)\pi_j.$$

我们强调, 如果 i 是瞬态, j 是常返状态, 则 P_{ij}^n 的极限既取决于 i 也取决于 j. 这和 i, j 都属于常返类的情形形成鲜明的对比.

例 (具有 $n+1$ 个状态的赌徒输光模型)

$$\begin{Vmatrix} 1 & 0 & 0 & 0 & \cdots & & \\ q & 0 & p & 0 & \cdots & & \\ 0 & q & 0 & p & \cdots & & \\ \vdots & & & & & & \\ & & \cdots & q & 0 & p \\ & & \cdots & 0 & 0 & 1 \end{Vmatrix}$$

我们要计算 $u_i = \pi_i(C_0)$ 和 $v_i = \pi_i(C_n)$, 即过程从状态 i 出发最后分别进入吸收 (所以是常返的) 状态 0 和 n 的概率. 方程组 (3.4) 变为

$$\begin{aligned} u_1 &= q + pu_2, \\ u_i &= qu_{i-1} + pu_{i+1}, \quad 2 \leqslant i \leqslant n-2, \\ u_{n-1} &= qu_{n-2}, \end{aligned} \tag{3.6}$$

这是有 $n-1$ 个未知数的 $n-1$ 个非齐次方程组. 设解的形式是 $u_r = x^r$, 代入 (3.6) 中间方程, 约去公因子, 得

$$px^2 + q = x.$$

92

此方程存在两个解: $x = 1$ 和 $x = q/p$. 因此, $u_r = A + B\,(q/p)^r, r = 1, 2, \cdots, n-1$, 满足式 (3.6) 中间方程, 其中 A, B 为任意常数. 现在我们确定 A 和 B 使其满足第一个和最后一个方程. (若 $q = p$, 则 $x = 1$ 是 $px^2 + q = x$ 的双重根, 这时需用 r 代替 $(q/p)^r$.) 在 $q \neq p$ 的情况下, 我们得到条件

$$A + B\frac{q}{p} = q + p(A + B\frac{q^2}{p^2})$$

及

$$A + B\left(\frac{q}{p}\right)^{n-1} = q\left(A + B\left(\frac{q}{p}\right)^{n-2}\right),$$

化简为

$$A = 1 - B$$

及

$$p^n A + q^n B = 0.$$

求此方程组的解, 得

$$A = \frac{q^n}{q^n - p^n}, \quad B = \frac{-p^n}{q^n - p^n}.$$

代入 u_r 的表达式

$$u_r = \frac{(q/p)^n - (q/p)^r}{(q/p)^n - 1}, \quad 若p \neq q.$$

如果 $q = p$, 我们类似可得 $A = 1, B = -\frac{1}{n}$, 于是

$$u_r = \frac{n-r}{n}, \quad 若p = q.$$

用类似方法可得

$$v_i = 1 - u_i,$$

这正是我们所要的结果, 因为一定被类 C_0 和 C_n 之一吸收是显然.

考虑对手拥有无限财富的赌徒输光模型. 赌徒输光 (吸收到 0) 的概率满足方程

$$\begin{aligned} u_1 &= q + pu_2, \\ u_i &= qu_{i-1} + pu_{i+1}, \quad i \geqslant 2. \end{aligned} \tag{3.7}$$

我们同样求得

$$u_i = A + B(q/p)^i, \quad q \neq p$$

和

93

$$u_i = A + Bi \quad (q = p = 1/2).$$

如果 $q \geqslant p$, 则由 u_i 有界知 $B = 0$, 由式 (3.7) 的第一个方程求得 $u_i = 1$. 如果 $q < p$, 我们可求得 $u_i = (q/p)^i$. 事实上, 由有限状态赌徒输光模型通过简单的极限过渡, 得 $\mu_1 = q/p$, 由此易推出 $u_i = (q/p)^i$.

3.4 常返性准则

我们证明两个定理, 它们对判断马尔可夫链的常返性或瞬时性是很实用的, 然后应用这两个定理研究几个例子.

定理 4.1 设一个不可约的马尔可夫链 $\mathfrak{P}$ 的状态空间为非负整数集. 则 $\mathfrak{P}$ 是瞬时的 (即每个状态是瞬时的) 充分必要条件为方程组

$$\sum_{j=0}^{\infty} P_{ij} y_j = y_i, \quad i \neq 0, \tag{4.1}$$

存在非常数有界解.

证明 设 $\mathfrak{P}$ 的转移矩阵为

$$\boldsymbol{P}=\|P_{ij}\|=\left\|\begin{matrix} P_{00} & P_{01} & \cdots \\ P_{10} & P_{11} & \cdots \\ \cdots & \cdots & \cdots \end{matrix}\right\|,$$

与其相联系的新转移矩阵为

$$\widetilde{\boldsymbol{P}}=\|\widetilde{P}_{ij}\|=\left\|\begin{matrix} 1 & 0 & 0 & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ P_{20} & P_{21} & P_{22} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{matrix}\right\|. \tag{4.2}$$

我们用 $\widetilde{\mathfrak{P}}$ 来表示由转移矩阵 (4.2) 决定的马尔可夫链, 它把原来马尔可夫链的状态 "0" 改为吸收状态而其他状态的转移概率并不改变.

为证必要性, 我们将假设过程 $\mathfrak{P}$ 是瞬时的, 然后找出方程组 (4.1) 的非常数有界解.

令 f_{i0}^* 表示已知初始状态 i. 经过有限步之后终于达到状态 0 的概率. 由于过程 $\mathfrak{P}$ 是瞬时的, 则对某个 $j\neq 0$ 有 $f_{j0}^*<1$, 否则状态 0 为常返的. (为证明这一点, 应当记住在不可约马尔可夫链中所有状态或同时是常返的, 或同时是非常返的.) 对于过程 $\widetilde{\mathfrak{P}}$, 显然对某些 $j\neq 0, \widetilde{\pi}_0(C_0)=1, \widetilde{\pi}_j(C_0)=f_{j0}^*<1$, 且对所有 i 成立, $\widetilde{\pi}_i(C_0)=\sum\limits_{j=0}^{\infty}\widetilde{P}_{ij}\widetilde{\pi}_j(C_0)$. 因此 $\widetilde{\pi}_i(C_0)=\sum\limits_{j=0}^{\infty}P_{ij}\widetilde{\pi}_j(C_0)$ 对于 $i\neq 0$ 成立, 这样, $y_i=\widetilde{\pi}_j(C_0)(j=0,1,2,\cdots)$ 就是所要求的非常数有界解.

[94]

现在, 我们假设方程组 (4.1) 存在有界解 $\{y_i\}$, 则

$$\sum_{j=0}^{\infty}\widetilde{P}_{ij}y_j=y_i,\quad i\geqslant 0,$$

通过迭代, 我们得到, 对所有的 $i\geqslant 0$ 和所有的 $n\geqslant 1$,

$$\sum_{j=0}^{\infty}\widetilde{P}_{ij}^n y_j=y_i.$$

如果链是常返的, 则

$$\lim_{n\to\infty}\widetilde{P}_{i0}^n=1,$$

因此

$$\sum_{j\neq 0}\widetilde{P}_{ij}^n y_j\leqslant M(1-\widetilde{P}_{i0}^n)\to 0,\quad \text{当}n\to\infty\text{时},$$

这里 M 是 $\{y_j\}$ 的上界. 因此

$$y_i = \sum_{j \neq 0} \widetilde{P}_{ij}^n y_j + \widetilde{P}_{i0}^n y_0 \to y_0.$$

因此, 对于所有的 $i, y_i = y_0$, 即 $\{y_i\}$ 是常数列.

定理 4.2 一个不可约马尔可夫链为常返的充分条件是存在一个序列 $\{y_i\}$ 使得

$$\sum_{j=0}^{\infty} P_{ij} y_j \leqslant y_i, \ i \neq 0 \ \text{且} \ y_i \to \infty. \tag{4.3}$$

证明 利用前一定理证明中的相同记号, 我们有

$$\sum_{j=0}^{\infty} \widetilde{P}_{ij} y_j \leqslant y_i, \quad \text{对所有 } i.$$

因为 $z_i = y_i + b$ 满足 (4.3), 我们可假设对所有的 $i \geqslant 0, y_i > 0$. 迭代前面的不等式, 我们有

95

$$\sum_{j=0}^{\infty} \widetilde{P}_{ij}^m y_j \leqslant y_i.$$

给定 $\varepsilon > 0$, 我们选取 $M(\varepsilon)$ 使当 $i \geqslant M(\varepsilon)$ 时 $\dfrac{1}{y_i} \leqslant \varepsilon$. 由于

$$\sum_{j=0}^{M-1} \widetilde{P}_{ij}^m y_j + \sum_{j=M}^{\infty} \widetilde{P}_{ij}^m y_j \leqslant y_i,$$

所以

$$\sum_{j=0}^{M-1} \widetilde{P}_{ij}^m y_j + \min_{r \geqslant M} \{y_r\} \sum_{j=M}^{\infty} \widetilde{P}_{ij}^m \leqslant y_i.$$

因为

$$\sum_{j=0}^{\infty} \widetilde{P}_{ij}^m = 1,$$

所以我们有

$$\sum_{j=0}^{M-1} \widetilde{P}_{ij}^m y_j + \min_{r \geqslant M} \{y_r\} \left(1 - \sum_{j=0}^{M-1} \widetilde{P}_{ij}^m \right) \leqslant y_i.$$

正如前面定理证明中所看到的,

$$\lim_{n \to \infty} \widetilde{P}_{ij}^n = 0, \quad \text{对于} j \neq 0.$$

这样, 当 $m \to \infty$ 时取极限, 我们得到对固定的 i,

$$\widetilde{\pi}_i(C_0)y_0 + \min_{r \geqslant M}\{y_r\}(1 - \widetilde{\pi}_i(C_0)) \leqslant y_i$$

或

$$1 - \widetilde{\pi}_i(C_0) \leqslant \frac{1}{\min\limits_{r \geqslant M}\{y_r\}}(y_i - \widetilde{\pi}_i(C_0)y_0) \leqslant \varepsilon K,$$

其中

$$K = y_i - \widetilde{\pi}_i(C_0)y_0.$$

由 ε 的任意性, 和 $\widetilde{\pi}_i(C_0) \leqslant 1$, 我们有 $\widetilde{\pi}_i(C_0) = 1$ 对每个 i 成立, 这就证明了原过程是常返的. ■

3.5 一个排队例子

现在我们来讨论 2.2 节例 C 的排队模型. 其转移矩阵是

$$\|P_{ij}\| = \left\| \begin{matrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & a_0 & a_1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{matrix} \right\|,$$

其中, $a_k > 0$ 且 $\sum\limits_{k=0}^{\infty} a_k = 1$.

96

(实际上, 在随后的分析中, 我们仅仅用到性质 $0 < a_0 < 1$ 和 $a_0 + a_1 < 1$, 它们保证了该马尔可夫链是不可约的.) 如果 $\sum\limits_{k=0}^{\infty} k a_k > 1$, 我们证明方程组 $\sum\limits_{j=0}^{\infty} P_{ij} y_j = y_i, i \neq 0$, 存在一个非常数有界解, 于是由定理 4.1 即可断言该过程是瞬时的. 令 $y_j = \xi^j$, 上述方程组取如下形式

$$\sum_{j=0}^{\infty} P_{ij}\xi^j = \sum_{j=i-1}^{\infty} a_{j-i+1}\xi^j = \xi^i$$

或

$$\sum_{j=i-1}^{\infty} a_{j-i+1}\xi^{j-i+1} = \xi = \sum_{k=0}^{\infty} a_k \xi^k = f(\xi), \quad (i \neq 0).$$

于是, $f(0) = a_0 > 0$ 且 $f(1) = \sum\limits_{k=0}^{\infty} a_k = 1$, 因此, 若 $f'(1) = \sum\limits_{k=0}^{\infty} k a_k > 1$, 则存在一个 $\xi_0, 0 < \xi_0 < 1$, 使 $f(\xi_0) = \xi_0$. 这由图 3-1 容易得到. 向量 $y_j = \xi_0^j, j = 0, 1, 2, \cdots$ 正是

所要的有界解, 且显然是非常数解. 如果 $\sum ka_k \leqslant 1$, 则应用定理 4.2, 其中 $y_j = j$, 如果 $i \neq 0$, 我们有

$$\begin{aligned}\sum_{j=0}^{\infty} P_{ij} j &= \sum_{j=i-1}^{\infty} a_{j-i+1} j \\ &= \sum_{j=i-1}^{\infty} a_{j-i+1}(j-i+1) + i - 1 \\ &= \sum_{k=0}^{\infty} k a_k - 1 + i \\ &\leqslant i.\end{aligned}$$

因此, 如果 $\sum ka_k \leqslant 1$, 则该过程是常返的.

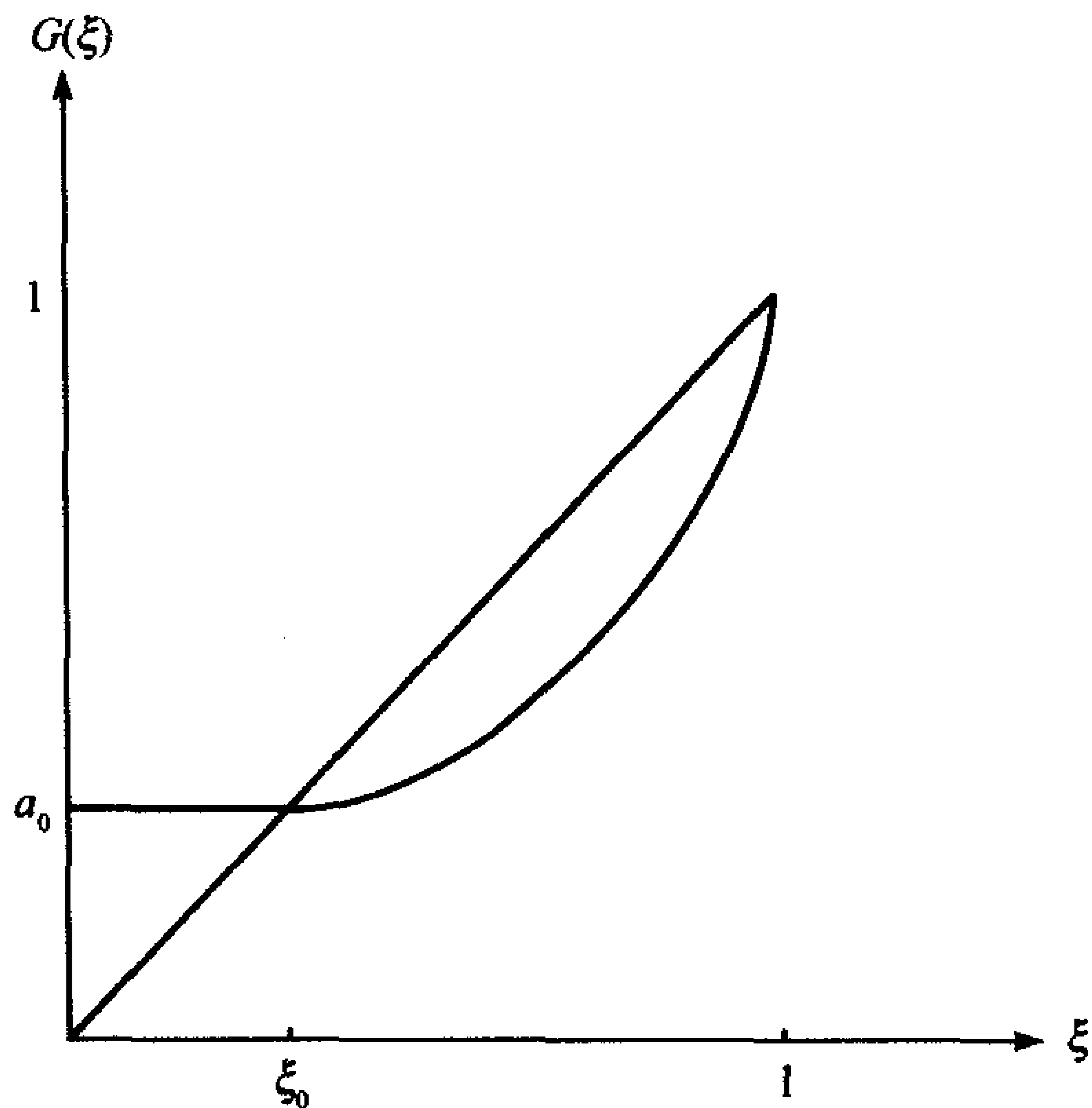


图 3-1

为了弄清过程 $\mathfrak{P}$ 究竟是零常返还是正常返, 我们先来讨论下面具有某种单独重要性的辅助问题.

设 $X_1, X_2, \cdots$ 表示取值于 $\{-1, 0, 1, 2, \cdots\}$ 的独立同分布随机变量序列, 且

$$\Pr\{X_i = k\} = b_k, \quad k = -1, 0, 1, 2, \cdots, \quad b_{-1} > 0,$$

并设 $S_n = X_1 + X_2 + \cdots + X_n$. 令 Z 是使 S_n 第一个变为负值的 n 值, 假定

$$\Pr\{Z = k\} = \gamma_k, \quad k = 1, 2, 3, \cdots. \tag{5.1}$$

令

$$U(s)=\sum_{k=0}^{\infty}\gamma_k s^k \quad (\gamma_0=0) \tag{5.2}$$

97

表示 (5.1) 的母函数. 若 $T_n^{(r)}=r+S_n$(r 是一个非负整数), 令 $Z^{(r)}$ 是一个随机变量, 其值等于使 $T_n^{(r)}<0$ 的第一个 n 值. 由于每个 $X_i\geqslant -1$, 我们容易验证 $Z^{(r)}=Z_1+Z_2+\cdots+Z_{r+1}$, 这里 Z_i 是以 (5.1) 为分布的独立同分布. 显然, $Z^{(r)}$ 的母函数是 $[U(s)]^{r+1}$. 令 $\gamma_m^{(r+1)}$ 表示 $U(s)^{r+1}$ 中 s^m 的系数.

最后, 记

$$G(s)=\frac{b_{-1}}{s}+b_0+b_1s+b_2s^2+\cdots.$$

我们的目的是利用 $G(s)$ 确定 $U(s)$. 为此先来推导通常的更新关系式:

$$\gamma_1=b_{-1},\quad \gamma_k=\sum_{j=0}^{\infty}b_j\gamma_{k-1}^{(j+1)},\quad k\geqslant 2. \tag{5.3}$$

上述关系式中第一个是显然的. 至于第二个, 事件 $\{S_n\geqslant 0, n=1,\cdots,k-1; S_k=-1\}$ 是不相交事件族 $\{X_1=j; X_2+\cdots+X_n+j\geqslant 0, n=2,\cdots,k-1; X_2+\cdots+X_k+j=-1\}, j=0,1,\cdots$ 的并, 它的概率显然等于 $b_j\gamma_{k-1}^{(j+1)}$, 因为 X_i 是独立同分布的. 利用全概率公式立即导出 (5.3). 由 (5.3) 过渡到母函数, 我们有

98

$$\begin{aligned}
U(s)&=b_{-1}s+\sum_{n=2}^{\infty}\left(\sum_{j=0}^{\infty}b_j\gamma_{n-1}^{(j+1)}\right)s^n\\
&=b_{-1}s+s\sum_{j=0}^{\infty}b_j\left(\sum_{n=2}^{\infty}\gamma_{n-1}^{(j+1)}s^{n-1}\right)\\
&=b_{-1}s+s\sum_{j=0}^{\infty}b_j[U(s)]^{j+1}\\
&=b_{-1}s+sU(s)\left[G(U(s))-\frac{b_{-1}}{U(s)}\right]\quad \text{对于} 0<s\leqslant 1\\
&=sU(s)G(U(s)).
\end{aligned}$$

由于 $U(s)$ 是连续的且当 $s\in[0,1]$ 时是严格增的, 同时 $U(0)=0$. 因此, 对于 $0<s\leqslant 1, U(s)$ 满足 $G(U(s))=\frac{1}{s}$. 但是对于 $s>0$,

$$G''(s)=\frac{2b_{-1}}{s^3}+2b_2+6b_3s+12b_4s^2+\cdots>0,$$

所以 $G(s)$ 是凸函数, 由 $G(s)$ 的定义推得 $\lim\limits_{s\downarrow 0}G(s)=+\infty$ 和 $G(1)=1$. 看一看下面图形 (图 3-2 和图 3-3) 便知方程 $G(x)=\frac{1}{s}$ 对于每个 $s\in[0,1]$ 至多有两个正解. 由

于 $\lim_{s\downarrow 0} U(s)=0$ 和 $U(s)$ 在 $[0,1]$ 上是严格增的, 我们可以看出, 如果 $G(x)=\frac{1}{s}$ 存在两个解, 则 $U(s)$ 必是 $G(x)=\frac{1}{s}$ 的两个解中较小的一个.

我们现在研究在什么条件下有 $\sum_{k=0}^{\infty}\gamma_k=1$ 或 <1. 可能出现下面两种情况:

情况 1: $G'(1)>0$.$G'(1)>0$ 等价于 $b_{-1}<\sum_{n=0}^{\infty}nb_n$, 而图 3-2 显然说明 $U(1)=\sum_{k=0}^{\infty}\gamma_k=\xi_0<1$. 所以事件 $\{S_n\geqslant 0$ 对于所有的 $n\}$ 的概率是严格正的.

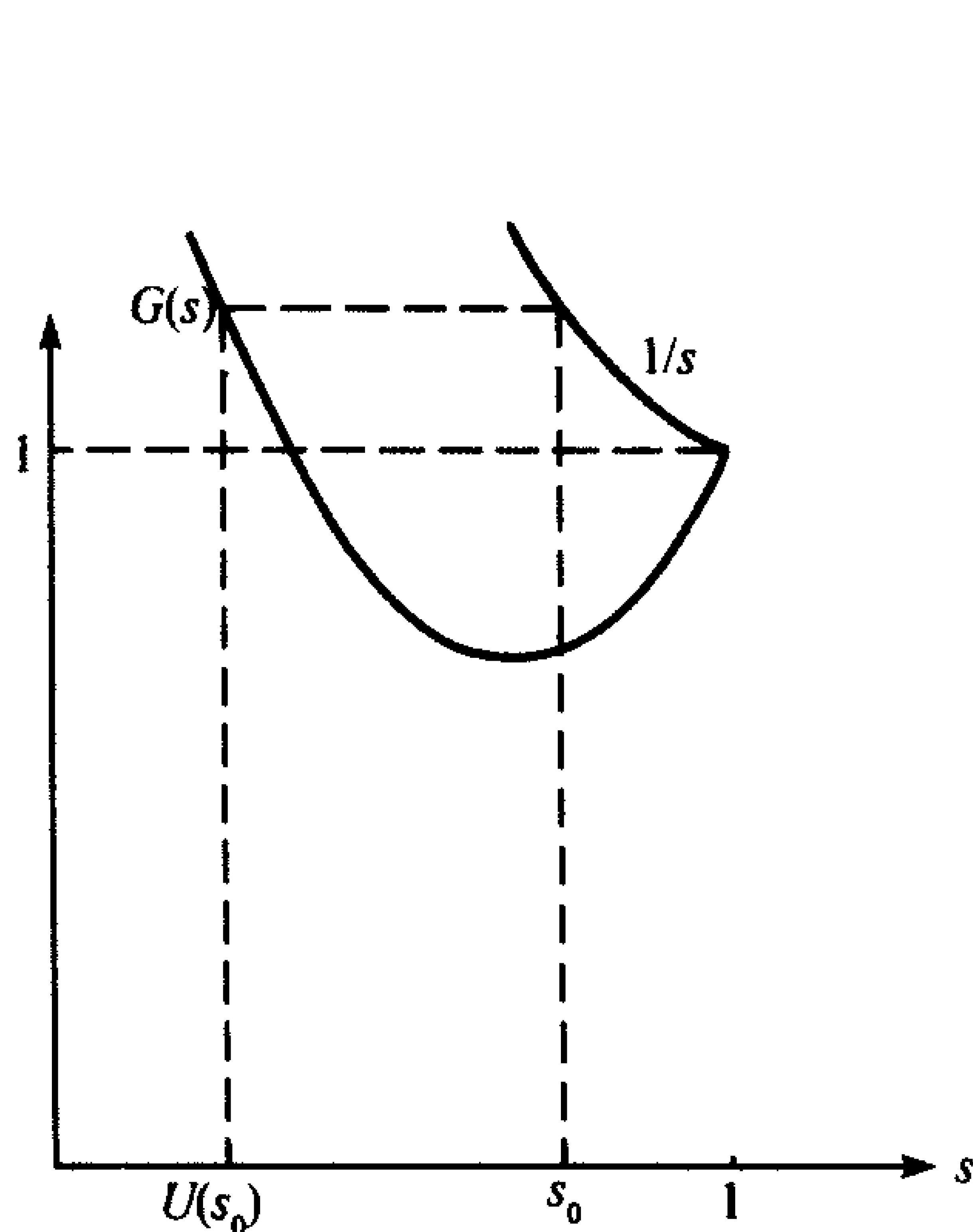


图 3-2

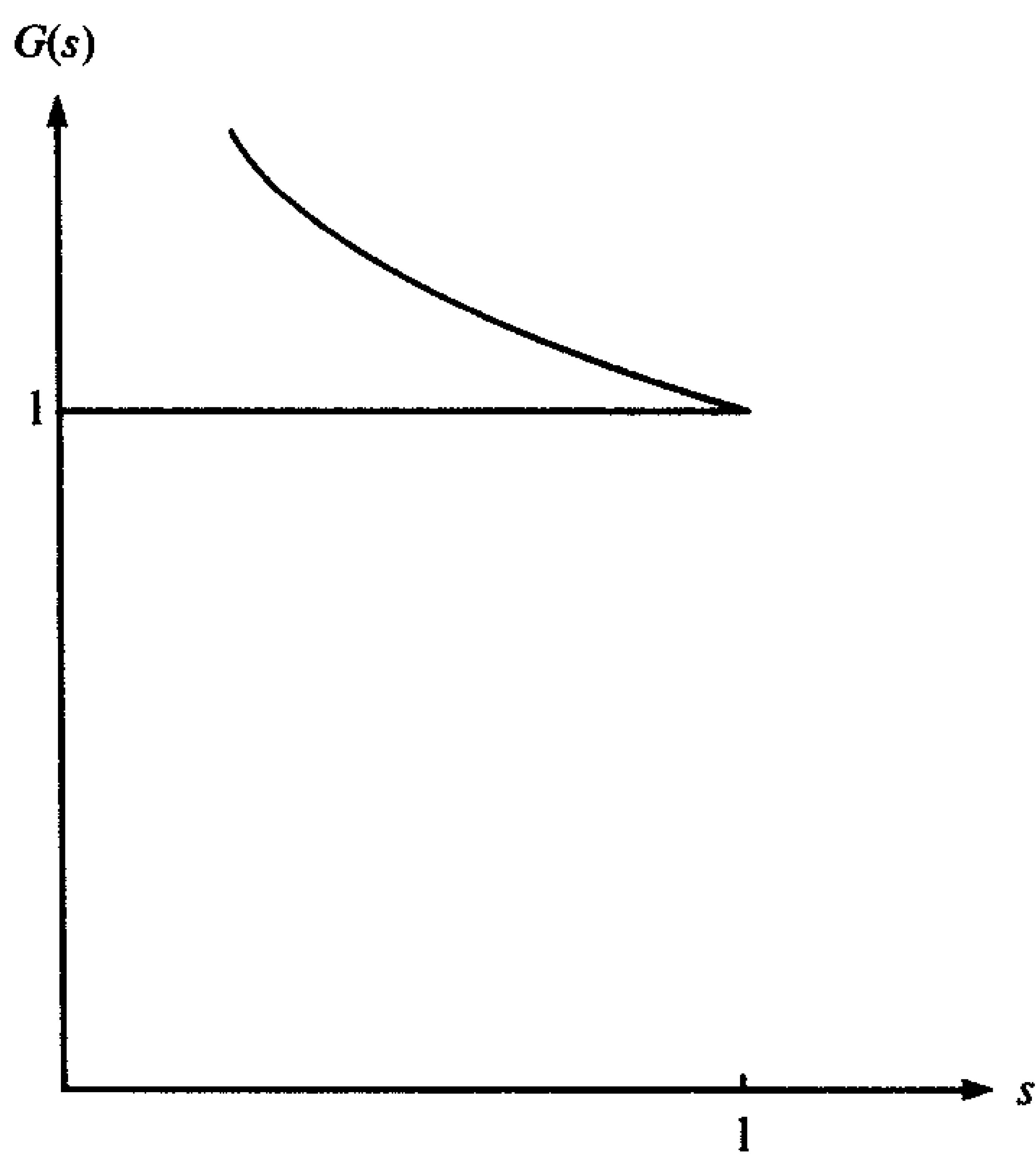


图 3-3

情况 2: $G'(1)\leqslant 0$.$G'(1)\leqslant 0$ 等价于 $b_{-1}\geqslant\sum_{n=0}^{\infty}nb_n$, 而这时我们有 $\sum_{k=0}^{\infty}\gamma_k=U(1)=1$. 因为对于 $0<s\leqslant 1, G'(U(s))U'(s)=-\frac{1}{s^2}$, 故在这种情况下当 $s\to 1$ 时 $U(s)\to 1$(见图 3-3). 这蕴涵着: 如果 $G'(1)<0$, 即如果

$$b_{-1}>\sum_{n=0}^{\infty}nb_n,$$

则

99

$$E(Z)=\sum_{n=0}^{\infty}n\gamma_n=U'(1)=\frac{-1}{G'(1)}<\infty,$$

如果 $G'(1)=0$, 即如果

$$b_{-1} = \sum_{n=0}^{\infty} n b_n,$$

则

$$E(Z) = \sum_{n=0}^{\infty} n \gamma_n = U'(1) = \infty.$$

100

现在我们再回到排队过程, 设 $a_k = b_{k-1}$, 且令 Z_{ij} = 从状态 i 出发第一次到达状态 $j < i$ 所需时间长度 (转移次数). 稍加思考便可发现, $Z_{i,i-1}$ 正是上述的随机变量 Z, 其母函数 $U(s)$ 已确定. 由于

$$\sum_{i=0}^{\infty} a_i = 1,$$

我们有

$$\left(b_{-1} > \sum_{n=0}^{\infty} n b_n\right) \longleftrightarrow \left(a_0 > \sum_{n=0}^{\infty} n a_{n+1}\right)$$
$$\longleftrightarrow \left(1 > \sum_{n=0}^{\infty} n a_n\right),$$

类似地有

$$\left(b_{-1} = \sum_{n=0}^{\infty} n b_n\right) \longleftrightarrow \left(a_0 = \sum_{n=0}^{\infty} n a_{n+1}\right)$$
$$\longleftrightarrow \left(1 = \sum_{n=0}^{\infty} n a_n\right).$$

因此, 如果 $\sum_{n=0}^{\infty} n a_n < 1$. 则 $E(Z_{i,i-1}) = \mu < \infty$, 如果 $\sum_{n=0}^{\infty} n a_n = 1$ 则 $E(Z_{i,i-1}) = \mu = \infty$. 由于一次只允许向后移动一步 (即过程此时是“持续”的), 我们有

$$Z_{ij} = Z_{i,i-1} + Z_{i-1,i-2} + \cdots + Z_{j+1,j}, j < i,$$

因此 $E(Z_{ij}) = (i-j)\mu$, 特别地有 $E(Z_{i,0}) = i\mu$.

现在让我们考虑状态 0 的平均返回时间. 首先注意到返回时间等于 1 的概率恰是 a_0(即转移概率 P_{00}). 此外, 从 0 出发经过两步或多步转移第一次回到 0 的样本函数可按照发生第一次转移所到达的状态 i 进行分组. 这样一组的平均返回时间正好等于 1 加上从状态 i 到达状态 0 所需要的平均时间. 这样的分解结合马尔可夫

性, 导出了下面平均返回时间的表达式:

$$\begin{aligned}\sum_{n=0}^{\infty} n f_{00}^{n} &= E(\text{返回时间}) \\ &= a_0 + \sum_{i=1}^{\infty} a_i \left[E(Z_{i,0}) + 1\right] = 1 + \sum_{i=1}^{\infty} a_i E(Z_{i,0}) \\ &= 1 + \sum_{i=1}^{\infty} i\mu a_i = 1 + \mu \sum_{i=0}^{\infty} i a_i.\end{aligned}$$

这样, 若 $\mu < \infty$, 即 $\sum_{i=0}^{\infty} i a_i < 1$, 则

101

$$\sum_{n=0}^{\infty} n f_{00}^{n} < \infty;$$

且, 若 $\mu = \infty$, 即 $\sum_{i=0}^{\infty} i a_i = 1$, 则

$$\sum_{n=0}^{\infty} n f_{00}^{n} = \infty.$$

综上所述, 我们有

$$\sum_{n=0}^{\infty} n a_n < 1 \ \Rightarrow \text{正常返},$$

$$\sum_{n=0}^{\infty} n a_n = 1 \ \Rightarrow \text{零常返},$$

以及

$$\sum_{n=0}^{\infty} n a_n > 1 \Rightarrow \text{瞬态}.$$

这些结论比较直观. 表达式 $\sum_{n=0}^{\infty} n a_n$ 是顾客在一个服务周期内平均到达数. 这样, 如果 $\sum_{n=0}^{\infty} n a_n > 1$, 那么在平均的意义上说, 每个周期内到达的人数比受服务的人要多, 因此我们可以料定等候队伍的长度将超出任何界限. 另一方面, 如果 $\sum_{n=0}^{\infty} n a_n < 1$, 过程的状态渐趋稳定. 然而平稳分布的计算是相当复杂的 (见第 18 章).

3.6 另一个排队模型

过程的状态是等候队伍的长度, 在每个单位时间里有一个人到达, 如果等候队伍至少有 k 个人, 则有 k 个人被服务的概率为 $a_k > 0, k = 0, 1, 2, \cdots$[1] 容易计算转移概率矩阵为

$$\|P_{ij}\| = \left\| \begin{array}{ccccc} \sum_{i=1}^{\infty} a_i & a_0 & 0 & 0 & \cdots \\ \sum_{i=2}^{\infty} a_i & a_1 & a_0 & 0 & \\ \sum_{i=3}^{\infty} a_i & a_2 & a_1 & a_0 & \\ \vdots & \vdots & \vdots & \vdots & \end{array} \right\|$$

102

我们证明, 如果 $\sum_{k=0}^{\infty} k a_k > 1$, 于是过程是正常返的, 则存在一个平稳分布. 平稳分布存在的原因是: 受到服务的顾客的平均数是 $\sum k a_k > 1$ 而同时只有一个新顾客到达.

考虑方程 $\sum_{i=0}^{\infty} \xi_i P_{ij} = \xi_j$ 且令 $\xi_i = \xi^i$, 则

$$\sum_{i=j-1}^{\infty} \xi^i a_{i-j+1} = \xi^j, \quad 对于 j \geqslant 1;$$

即

$$\sum_{i=j-1}^{\infty} \xi^{i-j+1} a_{i-j+1} = \xi,$$

通过变量替换化简为

$$\sum_{k=0}^{\infty} a_k \xi^k = \xi.$$

如果 $\xi (0 < \xi < 1)$ 满足这些方程, 则对 $j = 0$ 我们有

$$\begin{aligned} \sum_{i=0}^{\infty} P_{i0} \xi^i &= \sum_{i=0}^{\infty} \left(\sum_{k=i+1}^{\infty} a_k \right) \xi^i \\ &= \sum_{k=1}^{\infty} \sum_{i=0}^{k-1} a_k \xi^i (重新安排和的顺序) \end{aligned}$$

1. 准确地说应是, 当等候队伍恰有 k 人时, 则恰有 $i (i < k)$ 人被服务的概率为 a_i, 而 k 人被服务的概率为 $\sum_{i=k}^{\infty} a_i$.—— 译者注

$$= \sum_{k=1}^{\infty} a_k \left(\frac{1-\xi^k}{1-\xi} \right) = \frac{1}{1-\xi} \left(1 - a_0 - \sum_{k=1}^{\infty} a_k \xi^k \right)$$

$$= \frac{1}{1-\xi}(1 - a_0 - (\xi - a_0)) = 1,$$

所以方程对于 $j = 0$ 同样也满足.

我们考虑 $f(\xi) = \sum_{k=0}^{\infty} a_k \xi^k$. 由于 $f(0) = a_0 > 0$ 且 $f(1) = 1$, 如果 $f'(1) = \sum_{k=0}^{\infty} k a_k > 1$, 则存在 ξ_0 满足 $0 < \xi_0 < 1$ 且 $f(\xi_0) = \xi_0$(见图 3-4). 所以 $\pi_i = (1-\xi_0)\,\xi_0^i$ $(i = 0, 1, 2, \cdots)$ 是过程状态的平稳概率, 这里, $\sum_{i=0}^{\infty} \pi_i = 1$. 特别地, 从长远来看, 等候队伍空无一人的概率为 $1 - \xi_0$.

103

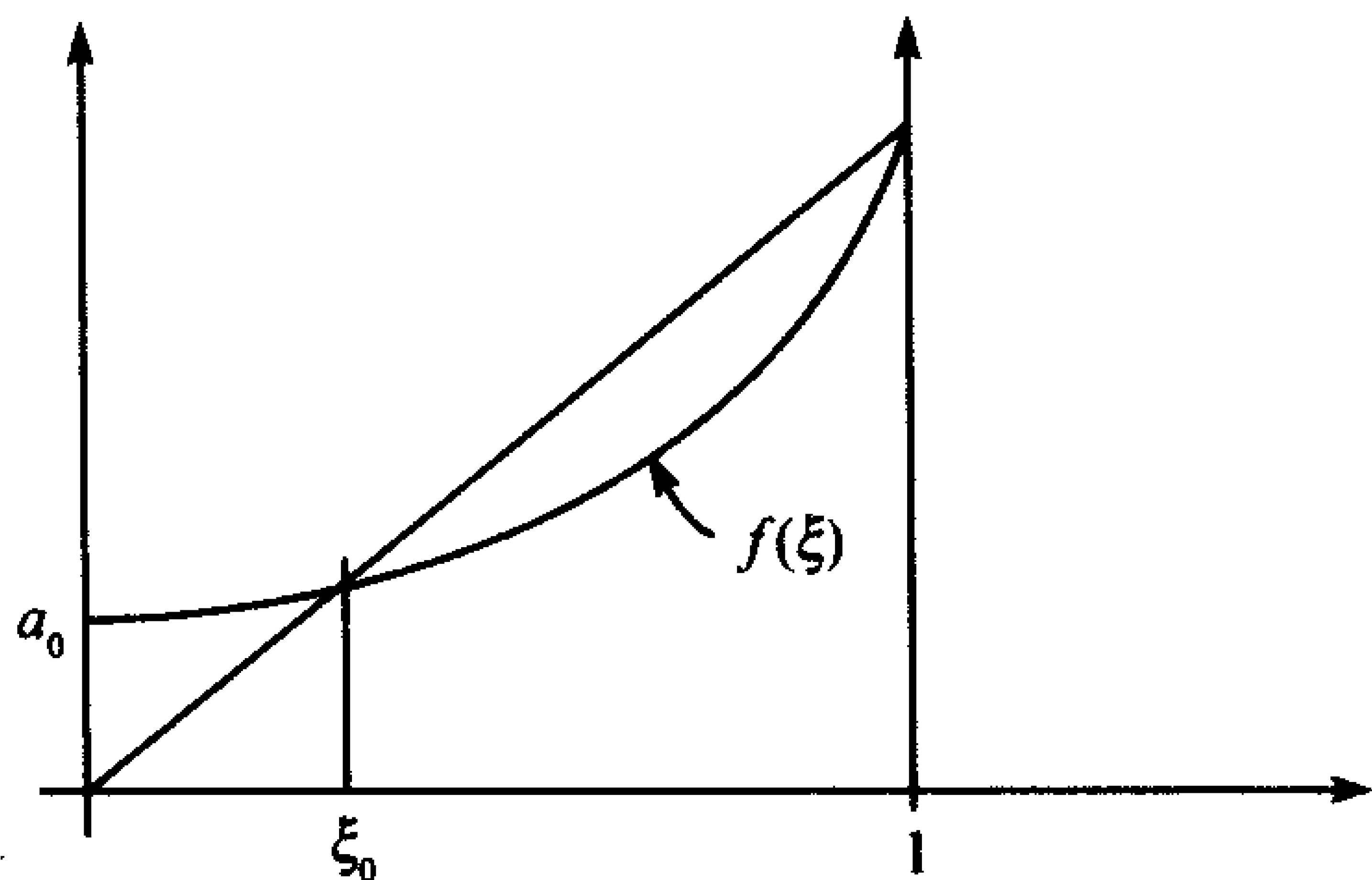


图 3-4

方程组 $\sum_{i=0}^{\infty} \xi_i P_{ij} = \xi_j, j \neq 0$, 等价与方程组 $\sum_{j=0}^{\infty} \widetilde{P}_{ij} \xi_j = \xi_i, i \neq 0$, 其中 $\widetilde{P}_{ij}$ 是 3.5 节中矩阵的元素. 在 $\sum_{k=0}^{\infty} k a_k \leqslant 1$ 的情况下, 过程 $\widetilde{P}$(如我们已见到的) 是常返的, 因此后一个方程组没有非常数有界解. 因此, 如果 $\sum k a_k \leqslant 1$, 方程组 $\sum_{i=0}^{\infty} \eta_i P_{ij} = \eta_j$ 没有有界解, 从而过程不存在平稳分布, 于是该过程或者是零常返的, 或者是非常返的. 我们现在证明方程组

$$\sum_{j=0}^{\infty} P_{ij} y_j = y_i, \quad i \neq 0 \tag{6.1}$$

有非常数有界解当且仅当 $\sum k a_k < 1$, 因此, 当且仅当 $\sum k a_k < 1$ 时过程是非常返的, 而当 $\sum_{k=0}^{\infty} k a_k = 1$ 时过程一定是零常返的. 因为 (6.1) 有一个常数解, 所以我们可

以令 $y_0 = 0$, 于是 (6.1) 化简为

$$
\begin{aligned}
a_2y_0 + a_1y_1 + a_0y_2 &= y_1 \\
a_3y_0 + a_2y_1 + a_1y_2 + a_0y_3 &= y_2 \\
\cdots\cdots\cdots\cdots\cdots \\
a_{n+1}y_0 + a_ny_1 + \cdots + a_1y_n + a_0y_{n+1} &= y_n \\
\cdots\cdots\cdots\cdots\cdots
\end{aligned}
$$

用 s^{i+1} 乘以第 i 个方程, 然后求和, 并令

$$Y(s) = \sum_{k=0}^{\infty} y_k s^k, \quad A(s) = \sum_{k=0}^{\infty} a_k s^k,$$

利用卷积公式, 我们得到

$$Y(s)A(s) - sa_0y_1 = sY(s) \tag{6.2}$$

若 $A(s) \neq s$, 由上式得

$$Y(s) = \frac{sa_0y_1}{A(s) - s}. \tag{6.2$'$}$$

如果 $A'(1) = \sum_{k=0}^{\infty} ka_k > 1$, 由于 $A(0) = a_0, A(1) = 1$, 则必存在某个 $s, 0 < s < 1$, 使得 $A(s) = s$. 因为 $Y(s)$ 对于每个 $s \in [0,1]$ 收敛, 所以 $Y(s)$ 没有有界系数, 从而当 $\sum_{k=0}^{\infty} ka_k > 1$ 时过程是常返的.

如果 $A'(1) = \sum_{k=0}^{\infty} ka_k \leqslant 1$, 由 $A(s)$ 的严格凸性, 即 $A''(s) > 0$, 推知对于 $0 < s < 1$ 均有 $A(s) \neq s$(见图 3-5). 考虑情况 $\sum ka_k \leqslant 1$,

104

$$
\begin{aligned}
A(s) - s &= (1-s)\left[1 - \frac{1 - A(s)}{1 - s}\right] \\
&= (1-s)\left[1 - (1 - A(s))\sum_{k=0}^{\infty} s^k\right] \\
&= (1-s)\left[1 - \sum_{n=0}^{\infty}\left(1 - \sum_{i=0}^{n} a_i\right) s^n\right] \\
&= (1-s)\left[1 - \sum_{n=0}^{\infty}\left(\sum_{i=n+1}^{\infty} a_i\right) s^n\right] \quad \text{(重排和的顺序)} \\
&= (1-s)\left[1 - W(s)\right], \quad W(s) = \sum_{n=0}^{\infty} w_n s^n,
\end{aligned}
$$

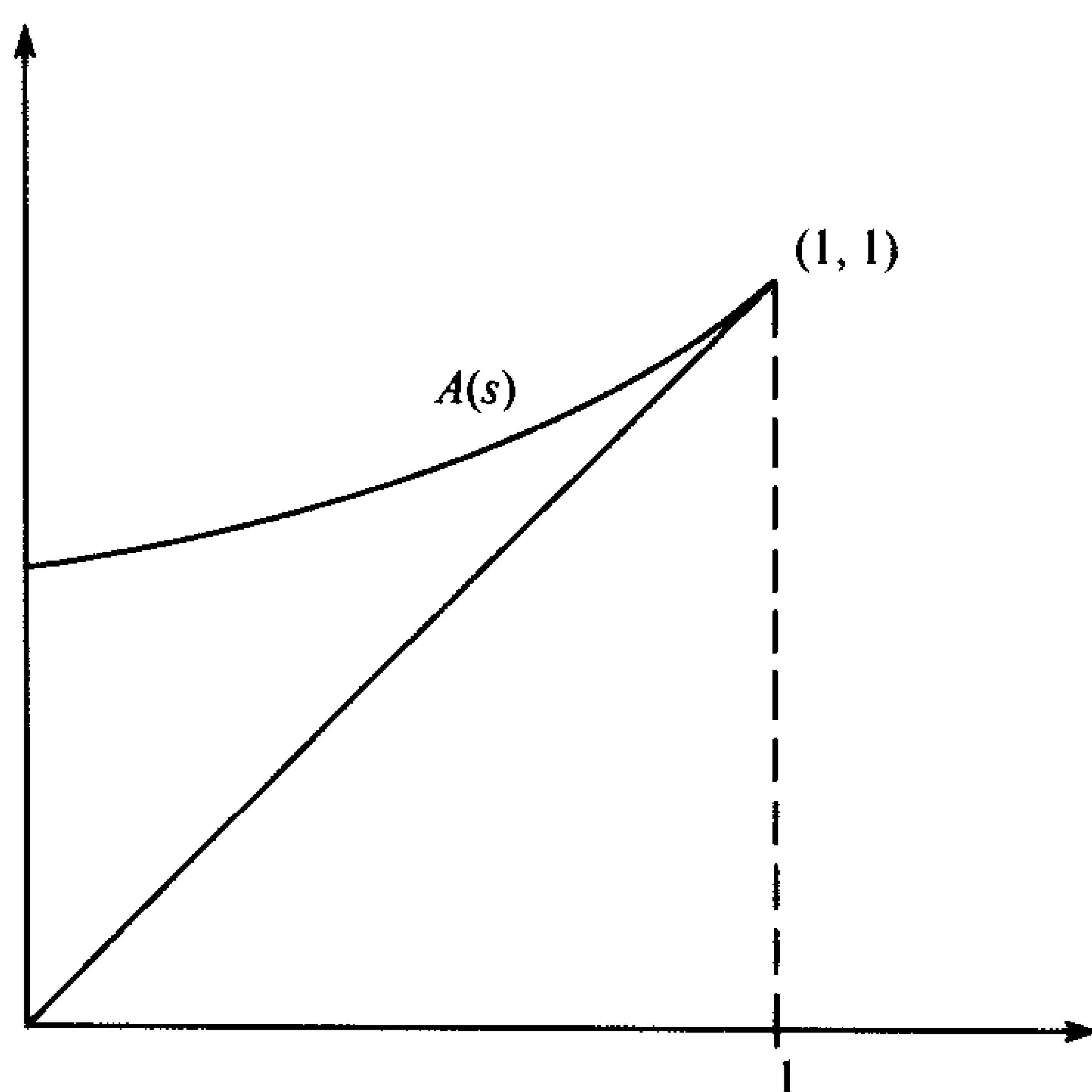


图 3-5

其中

$$w_n = \sum_{i=n+1}^{\infty} a_i > 0,$$

且

105

$$\sum_{n=0}^{\infty} w_n = \sum_{n=0}^{\infty} \left(\sum_{i=n+1}^{\infty} a_i \right) = \sum_{k=0}^{\infty} k a_k \leqslant 1.$$

故

$$\begin{aligned} Y(s) &= \frac{s a_0 y_1}{(1-s)[1-W(s)]} \\ &= \frac{s a_0 y_1}{1-s}(1 + W(s) + (W(s))^2 + \cdots) \\ &= s a_0 y_1 \frac{U(s)}{1-s}, \quad \text{其中} \quad U(s) = \sum_{n=0}^{\infty} u_n s^n = \sum_{k=0}^{\infty} [W(s)]^k, u_n \geqslant 0. \\ &= s a_0 y_1 V(s), \end{aligned}$$

其中

$$V(s) = \sum\nolimits_{n=0}^{\infty} v_n s^n, \quad v_n = \sum\nolimits_{k=0}^{n} u_k.$$

于是

$$V(s) = \frac{U(s)}{1-s},$$

由于 $U(1)=1+W(1)+(W(1))^2+\cdots$ 是收敛几何级数, 我们有

$$\left(W(1)=\sum ka_k<1\right)\longleftrightarrow\left(U(1)=\sum_{k=0}^{\infty}u_k<\infty\right).$$

显然, $v_1<v_2<\cdots\to U(1)$, 所以, 当且仅当 $\sum ka_k<1$ 时 $Y(s)=sa_0y_1V(s)$ 幂级数展开具有有界系数. 所以, 如果 $\sum ka_k<1$, 取 $y_1\neq 0, y_k=a_0y_1v_{k-1}$, 依方程 (6.2) 并令其系数相等, 我们得到 (6.1) 的一个有界非常数解. 由此推得该过程是非常返的. 如果 $\sum ka_k=1$, 则 (6.1) 的任何解一定无界, 从而该过程是常返的. 综述如下:

若 $\sum ka_k<1$, 过程是非常返的;

若 $\sum ka_k=1$, 过程是零常返的;

若 $\sum ka_k>1$, 过程是正常返的.

3.7 随机游动

我们应用 3.4 节的常返准则于随机游动, 其转移矩阵为

$$\|P_{ij}\|=\begin{Vmatrix} r_0 & p_0 & 0 & 0 & \cdots \\ q_1 & r_1 & p_1 & 0 & \cdots \\ 0 & q_2 & r_2 & p_2 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \end{Vmatrix}.$$

令 [106]

$$\pi_0=1,\quad \pi_n=\frac{p_0p_1\cdots p_{n-1}}{q_1q_2\cdots q_n}.$$

对于情况 $r_i\equiv 0$, 我们已证明 (见 3.1 节例) 当且仅当 $\sum_{n=0}^{\infty}\pi_n<\infty$ 时上述随机游动具有平稳分布. 现在我们考虑方程组

$$\sum_{j=0}^{\infty}P_{ij}y_j=y_i,\quad i\neq 0,$$

或

$$\begin{gathered} q_1y_0+r_1y_1+p_1y_2=y_1, \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ q_ny_{n-1}+r_ny_n+p_ny_{n+1}=y_n. \\ \cdots\cdots\cdots\cdots\cdots\cdots \end{gathered}$$

通过检验表明其解张成一个二维线性空间. 我们可以事先任意指定 y_0 和 y_1, 则所有其他 y_i 可由这些方程确定. $y_i \equiv 1$ 是方程组的平凡解. 我们证明, $y_0 = 0, y_n = \sum_{i=0}^{n-1} \frac{1}{p_i\pi_i}, n \geqslant 1$, 也是一个解. 对于第一个方程

$$q_1 y_0 + r_1 y_1 + p_1 y_2 = r_1\left(\frac{1}{p_0}\right) + p_1\left(\frac{1}{p_0} + \frac{q_1}{p_1 p_0}\right) = \frac{1}{p_0} = y_1.$$

对于第 n 个方程, 我们必须证明

$$q_n\left(\sum_{i=0}^{n-2} \frac{1}{p_i\pi_i}\right) + r_n \sum_{i=0}^{n-1} \frac{1}{p_i\pi_i} + p_n \sum_{i=0}^{n} \frac{1}{p_i\pi_i} = \sum_{i=0}^{n-1} \frac{1}{p_i\pi_i}.$$

由于 $p_n + r_n + q_n = 1$, 只须验证下式

$$q_n \sum_{i=0}^{n-2} \frac{1}{p_i\pi_i} + p_n \sum_{i=0}^{n} \frac{1}{p_i\pi_i} = (p_n + q_n) \sum_{i=0}^{n-1} \frac{1}{p_i\pi_i}.$$

但左端正是

$$(q_n + p_n) \sum_{i=0}^{n-1} \frac{1}{p_i\pi_i} - q_n \frac{1}{p_{n-1}\pi_{n-1}} + p_n \frac{1}{p_n\pi_n},$$

而据 π_n 的定义,

$$-q_n \frac{1}{p_{n-1}\pi_{n-1}} = \frac{-1}{(p_{n-1}/q_n)\pi_{n-1}} = \frac{-1}{\pi_n},$$

[107] 即证得结论. 由于两个解 $(y_i \equiv 1)$ 和 $\left(y_n = \sum_{i=1}^{n-1} \frac{1}{p_i\pi_i}\right)$ 显然是独立的, 所以一般解可以表示为 $z_n = \alpha + \beta y_n$, 并且当且仅当 y_n 有界即 $\sum_{i=0}^{\infty} \frac{1}{p_i\pi_i} < \infty$ 时, $\sum_{j=0}^{\infty} P_{ij} z_j = z_i, i \neq 0$, 存在一个非常数有界解.

所以, 我们有

$$\sum_{i=0}^{\infty} \frac{1}{p_i\pi_i} = \infty \Longrightarrow \text{常返},$$

$$\sum_{i=0}^{\infty} \frac{1}{p_i\pi_i} = \infty \quad \text{且} \quad \sum_{i=0}^{\infty} \pi_i = \infty \Longrightarrow \text{零常返},$$

$$\sum_{i=0}^{\infty} \frac{1}{p_i\pi_i} = \infty \quad \text{且} \quad \sum_{i=0}^{\infty} \pi_i < \infty \Longrightarrow \text{正常返},$$

$$\sum_{i=0}^{\infty} \frac{1}{p_i\pi_i} < \infty \Longrightarrow \text{非常返}.$$

初等问题

1. 一个矩阵 $\boldsymbol{P}=\|P_{ij}\|_{i,j=1}^{\infty}$ 称为随机的, 如果

(i) $P_{ij}\geqslant 0$ 对所有 $i,j=1,2,\cdots$;

(ii) $\sum\limits_{j=1}^{\infty}P_{ij}=1$ 对所有 $i=1,2,\cdots$.

一个矩阵 $\boldsymbol{P}$ 称为双重随机的, 如果除满足条件 (i), (ii) 之外, 还满足

$$\sum_{i=1}^{\infty}P_{ij}=1 \quad \text{对所有} \quad j=1,2,\cdots.$$

证明: 若一个有限不可约马尔可夫链具有双重随机转移概率矩阵, 则所有平稳概率相等.

2. 状态空间为 $\{0,1,2,3,4,5\}$ 的马尔可夫链具有转移概率矩阵

$$\text{(a)}\begin{bmatrix}\frac{1}{3}&\frac{2}{3}&0&0&0&0\\ \frac{2}{3}&\frac{1}{3}&0&0&0&0\\ 0&0&\frac{1}{4}&\frac{3}{4}&0&0\\ 0&0&\frac{1}{5}&\frac{4}{5}&0&0\\ \frac{1}{4}&0&\frac{1}{4}&0&\frac{1}{4}&\frac{1}{4}\\ \frac{1}{6}&\frac{1}{6}&\frac{1}{6}&\frac{1}{6}&\frac{1}{6}&\frac{1}{6}\end{bmatrix},\qquad \text{(b)}\begin{bmatrix}1&0&0&0&0&0\\ 0&\frac{3}{4}&\frac{1}{4}&0&0&0\\ 0&\frac{1}{8}&\frac{7}{8}&0&0&0\\ \frac{1}{4}&\frac{1}{4}&0&\frac{1}{8}&\frac{3}{8}&0\\ \frac{1}{3}&0&\frac{1}{6}&\frac{1}{6}&\frac{1}{3}&0\\ 0&0&0&0&0&1\end{bmatrix},$$

试找出所有的类并计算极限概率 $\lim\limits_{n\to\infty}P_{5i}^{n}(i=0,1,2,3,4,5)$.

3. 考虑赌徒输光模型, 其中赌徒 I 的初始财富为 $a(a>10)$, 赌徒 II 的初始财富为 $b(b>10)$, 设 p 是赌徒 I 每局从 II 赢得一个单位财富的概率. 求赌徒 I 的财富在减少到 5 之前达到 $a+b-3$ 的概率.

108

4. 设 Y_n 是一均匀骰子 n 次独立滚动后出现的点数之和. 求

$$\lim_{n\to\infty}\Pr\{Y_n\text{是 13 的倍数}\}.$$

5. 考虑一马尔可夫链, 其转移矩阵

$$\boldsymbol{P}=\begin{bmatrix}p_0&p_1&p_2&\cdots&p_m\\ p_m&p_0&p_1&\cdots&p_{m-1}\\ \vdots&\vdots&\vdots&&\vdots\\ p_1&p_2&p_3&\cdots&p_0\end{bmatrix},$$

其中 $0 < p_0 < 1$ 且 $p_0 + p_1 + \cdots + p_m = 1$. 试确定其平稳分布 $\lim\limits_{n\to\infty} P_{ij}^n$.

6. 设某群体具有无限多成员, 其成员对某疾病或具免疫力或易感染. 令 X_n 表示第 n 段时间该群体中易感染此病的成员数目, 并假设 $X_0 = 0$ 且在该疾病不流行的情况下有 $X_{n+1} = X_n + 1$. 这样, 在此疾病不发生的情况下群体中易感染者的数目随时间增加可能是由于个体失去免疫力, 或者是一个新的易感染者进入该群体引起的.

但在每段时间内都存在发生此传染病的可能性, 发生概率为未知常数 p. 当疾病出现时, 所有易感染者都患此病, 此病不致命且产生免疫力, 所以若 T 是疾病出现的首次时间, 则 $X_T = 0$.

试计算 (X_n) 的平稳分布.

7. 一个飞机票预定系统由两台计算机组成, 在任一时刻都只有一台计算机在工作, 且在任一天计算机失灵的概率为 p. 有一套维修设备, 需要两天时间才能使失灵的计算机修复. 此维修设备一次只能修理一台计算机. 取数对 (x, y) 作为过程的状态, 其中, x 是在一天末可以运转的计算机台数, 经过一天修理尚未修复失灵计算机时, $y = 1$, 其余情况, $y = 0$. 此时构成状态空间为数对集的马尔可夫链, 其转移概率矩阵为:

出发状态 \ 到达状态→	(2,0)	(1,0)	(1,1)	(0,1)
(2,0)	q	p	0	0
(1,0)	0	0	q	p
(1,1)	q	p	0	0
(0,1)	0	1	0	0

$$\boldsymbol{P} = \begin{pmatrix} q & p & 0 & 0 \\ 0 & 0 & q & p \\ q & p & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

109 其中 $p + q = 1$. 试利用 p 和 q 求过程的平稳分布.

8. 考虑一生产线, 其中每个产品是废品的概率为 p. 假设每个产品的情况 (废品或正品) 与其他产品的情况无关. 现使用以下的抽样方法:

开始时抽取生产的每个产品, 直到连续取到 i 个正品, 然后从每 r 个产品中随机抽取一个, 直到发现一个废品, 当此情况发生后, 又恢复原来方法抽取生产出来的每个产品, 直到出现连续 i 个正品为止, 等等.

状态 $E_k(k = 0, 1, \cdots, i)$ 表示在 100% 抽样阶段发现 k 个连续正品, 而状态 E_{i+1} 表示在第二阶段抽样中 (即从 r 个产品中抽取一个) 至少有一个正品被抽取. (我们把第 m 次观测看作时刻 m.) 则状态序列是一个马尔可夫链, 其转移概率为

$$P_{jk} = \Pr\{\text{第 } m+1 \text{ 次观测后处于状态 } E_k \mid \text{在第 } m \text{ 次观测后处于状态} E_j\}$$

$$= \begin{cases} p & \text{若} \quad k = 0, j = 0, 1, \cdots, i, i+1. \\ 1 - p & \text{若} \quad k = j+1, j = 0, 1, \cdots, i, \text{或} \quad k = j = i+1. \\ 0 & \text{其他}. \end{cases}$$

对于所有 m.

(a) 试确定平稳分布.

(b) 试确定长远之后检验过的产品所占的比例.

(c) 试确定平均消耗量 (AOQ), 即在长远之后抽样产品中废品所占的比例.

9. 社会学家常假设在一个家庭中相继各代的社会等级可看作一个马尔可夫链. 这样, 儿子的职业仅与其父亲的职业有关而与祖父无关. 假如这样的模型是适当的, 并且转移概率矩阵由下面给出:

		儿子的等级		
		低	中	高
父亲的等级	低	0.40	0.50	0.10
	中	0.05	0.70	0.25
	高	0.05	0.50	0.45

对这样一个群体, 长远之后处于中间等级的人所占比例是多少?

10. 假如任意一天的天气与前两天的天气有关, 准确地说, 如果今天和昨天都是晴天, 明天将以概率 0.8 为晴天; 如果今天是晴天而昨天是阴天, 则明天将以概率 0.6 为晴天; 如果今天是阴天而昨天是晴天, 则明天将以概率 0.4 为晴天; 如果今天和昨天均为阴天, 则明天以概率 0.1 为晴天. 110

倘若我们认为任一时刻的状态由当天和前一天的天气情况所确定, 上述模型可转化为马尔可夫链. 我们说过程处于

状态(s,s), 如果昨天和今天晴天,

状态(s,c), 如果昨天晴而今天阴,

状态(c,s), 如果昨天阴而今天晴,

状态(c,c), 如果昨天和今天都是阴天.

则转移概率矩阵是

		今天的状态			
		(s,s)	(s,c)	(c,s)	(c,c)
昨天的状态	(s,s)	0.8	0.2		
	(s,c)			0.4	0.6
	(c,s)	0.6	0.4		
	(c,c)			0.1	0.9

.

(a) 求出马尔可夫链的平稳分布.

(b) 长远之后, 晴天所占比例为多少?

11. 考虑一正 $2r+1$ 边形, 其顶点为 $V_1, V_2, \cdots, V_{2r+1}$. 设在每一点 V_k 上有非负质量 w_k^1, 且 $w_1^1 + w_2^1 + \cdots + w_{2r+1}^1 = 1$, 现在每点 V_k 上置以新质量 w_k^2, 它是相邻两点旧质量的代数平均, 即

$$w_k^2 = \frac{1}{2}(w_{k-1}^1 + w_{k+1}^1).$$

如此类似地做 n 次变换. 试确定 $\lim_{n\to\infty} w_k^n$.

答案: $1/(2r+1)$, 与 k 无关.

12. 假设用离散单位测量的灯泡寿命是随机变量 X, 分布律为 $\Pr\{X = k\} = p_k, (k = 1, 2, \cdots)$. 若从新灯泡开始, 当其损坏时立即换上新的灯泡, 至时刻 n 为止被替换灯泡数的期望值记为 u_n, 则 u_n 满足方程.

$$u_n = F_X(n) + \sum_{k=1}^{n} p_k u_{n-k}, \quad n = 1, 2, \cdots,$$

其中 $F_X(n) = \sum_{k \leqslant n} p_k$.

在一座大楼中, 出于经济原因, 通常一次替换所有灯泡 (已损坏或没有损坏的) 比单独替换一个灯泡更加便宜. 所谓"整批替换策略"是由周期 N 所确定, 它要求在时间 $0, 1, \cdots, N-1$ 期间替换已损坏的灯泡, 随后在时间 N 替换所有的灯泡 (已损坏的和未损坏的). 若 C_1 是整批替换中每个灯泡所需费用, C_2 是损坏替换中每个灯泡所需费用, 可以证明, 采取这样策略长远之后, 每个灯泡每次替换平均费用为 $[C_1 + C_2 u_{N-1}]/N$, 即一个替换周期平均费用除以周期时间长度. (这个结果的证明见第五章).

111

(a) 由直观注意到, 至时刻 n 为止平均更新次数 u_n 不收敛而无限变大. 关于 u_n 的更新方程不满足定理 1.1 中的什么条件?

(b) 试导出关于 v_n 的更新方程, 这里 $v_n = \Pr\{$ 在时刻 n 需要一次替换 $\}$, 因而 $v_n = u_n - u_{n-1}, n = 1, 2, \cdots, (u_0 = 0)$.

(c) 若 $p_1 = 0.4, p_2 = 0.3, p_3 = 0.2$, 而 $p_4 = 0.1$, 对于 $n = 1, 2, \cdots, 10$, 计算 v_n 和 $u_n = v_1 + v_2 + \cdots + v_n$.

(d) 若 $C_1 = 1$ 元, $C_2 = 2$ 元, 试计算 N 为何值时费用最小?

答案: (b) $v_n = p_n + \sum_{k=1}^{n} p_{n-k} v_k, \quad v_0 = p_0 = 0.$

(c) $v_1 = 0.4000 \qquad v_2 = 0.4600$

$v_3 = 0.5040 \qquad v_4 = 0.5196$

$v_5 = 0.4910 \qquad v_6 = 0.4991$

$v_7 = 0.5013 \qquad v_8 = 0.5005$

$v_9 = 0.4994 \qquad v_{10} = 0.5002$

(d) $N^* = 2$.

问　题

1. 考虑下面随机游动:

$$P_{i,i+1} = p, \qquad 0 < p < 1,$$
$$P_{i,i-1} = q = 1-p, \qquad i = 1, 2, \cdots, r-1,$$
$$P_{0,0} = P_{r,r} = 1.$$

求 $d(k) = E[$ 吸收至状态 0 或 r 的时间 $|$ 初始状态为 $k]$.

答案:

$$d(k) = \frac{k}{q-p} - \frac{r}{q-p} \cdot \frac{1-(q/p)^k}{1-(q/p)^r}, \qquad 若 \quad p \neq \frac{1}{2},$$
$$= k(r-k), \qquad 若 \quad p = \frac{1}{2}.$$

2. 设 $\boldsymbol{P} = \|P_{ij}\|$ 是不可约马尔可夫链转移概率矩阵且 $\boldsymbol{P}$ 是幂等的 (即 $\boldsymbol{P}^2 = \boldsymbol{P}$). 证明, 对于所有 i 和 j 有 $P_{ij} = P_{jj}$, 且此马尔可夫链是非周期的.

提示: 对平均值 $\dfrac{1}{n}\sum\limits_{m=1}^{n} P_{ij}^m$ 利用定理 1.2.

3. 考虑状态空间为 $\{0, 1, 2, \cdots, N\}$ 且转移概率矩阵为 $\boldsymbol{P} = \|P_{ij}\|_{i,j=0}^N$ 的有限马尔可夫链 $\mathfrak{M}$, 设其状态分为三个等价类: $\{0\}, \{1, 2, \cdots, N-1\}, \{N\}$, 且 0 和 N 是吸收状态, 均可由状态 $k = 1, 2, \cdots, N-1$ 达到, 而 $\{1, 2, \cdots, N-1\}$ 为瞬态类. 令状态 k 满足 $0 < k < N$. 我们定义一称为“回复过程”的辅助过程 $\widetilde{\mathfrak{M}}$, 其转移概率矩阵是把 $\|P_{ij}\|$ 中第一行和最后一行改变为 $\widetilde{P}_{0,k} = \widetilde{P}_{N,k} = 1$, 而其他行不变. 此回复过程显然不可约. 试证明, 过程 $\mathfrak{M}$ 从初始状态 k 出发至吸收状态的时间期望 u_k 等于 $\dfrac{1}{\pi_0 + \pi_N} - 1$, 其中 $\pi_0 + \pi_N$ 是过程 $\widetilde{\mathfrak{M}}$ 处于状态 0 或 N 的平稳概率.

112

提示: 利用状态的平稳概率与平均返回时间之间的关系.

4. 考虑状态为 $0, 1, \cdots, N$ 的离散时间马尔可夫链, 其转移概率为

$$P_{ij} = \begin{cases} \mu_i, & j = i-1, \\ \lambda_i, & j = i+1, \\ 1 - \lambda_i - \mu_i, & j = i, \\ 0, & |j-i| > 1. \end{cases} \qquad i, j = 0, 1, \cdots, N.$$

假设 $\mu_0 = \lambda_0 = \mu_N = \lambda_N = 0$ 且所有其他 μ_i 和 λ_i 均为正的, 过程初始状态是 k. 试确定 0 和 N 的吸收概率.

答案: 定义 $p_0 = 1, \quad p_i = \dfrac{\mu_1 \mu_2 \cdots \mu_i}{\lambda_1 \lambda_2 \cdots \lambda_i}$;

$$\Pr\{吸收至\ 0\} = 1 - \Pr\{吸收至\ N\}$$

$$= \frac{\sum_{i=k}^{N-1} p_i}{\sum_{i=0}^{N-1} p_i}.$$

5. 在问题 4 的条件下, 试确定平均吸收时间.

6. 考虑有 $N+1$ 个状态 $0,1,\cdots,N$ 的马尔可夫链, 其转移概率为

$$P_{ij} = \binom{N}{j} \pi_i^j (1-\pi_i)^{N-j}, \qquad 0 \leqslant i,j \leqslant N,$$

$$\pi_i = \frac{1-\mathrm{e}^{-2ai/N}}{1-\mathrm{e}^{-2a}}, \qquad a > 0.$$

注意, 此时 0 和 N 是吸收状态. 试验证 $\exp(-2aX_t)$ 是一个鞅 (或等价地, 证明恒等式 $E[\exp(-2aX_{t+1})|X_t] = \exp(-2aX_t)$), 其中 X_t 是时刻 $t(t=0,1,2,\cdots)$ 的状态. 利用此性质证明, 从状态 k 出发被状态 N 吸收的概率 $P_N(k)$ 为

113

$$P_N(k) = \frac{1-\mathrm{e}^{-2ak}}{1-\mathrm{e}^{-2aN}}.$$

提示: 利用在有限时间内必定出现状态被 0 或 N 吸收的事实以及关系式 (试验证之)

$$E[\exp(-2aX_0)] = E[\exp(-2aX_n)]$$

$$= P_N(k)\exp(-2aN) + (1-P_N(k)).$$

7. 考虑一有限群体 (所含个体总数固定为 N), 其个体或为 A 型或为 a 型. 设在时刻 $t_1 < t_2 < t_3 < \cdots$, 某个体死亡且被另一个体 (A 或 a 型) 所代替. 若在时刻 t_n 替换前存在 j 个 A 型个体和 $N-j$ 个 a 型个体, 假设其中一个 A 型个体死亡概率是 $j\mu_1/B_j$, 一个 a 型个体死亡概率是 $(N-j)\mu_2/B_j$, 这里 $B_j = \mu_1 j + \mu_2(N-j)$. 这个模型基于如下想法: 一般地, 一个 A 型个体在时刻 t_n 死亡的可能性为 $\mu_1/(\mu_1+\mu_2)$, 而一个 a 型个体在时刻 t_n 死亡的可能性为 $\mu_2/(\mu_1+\mu_2)$. (μ_1/μ_2 可解释为 A 型相对于 a 型的选择优势). 考虑到群体的大小, 我们指定替换 A 型个体的概率为 $j\mu_1/B_j$ 替换 a 型个体的概率为 $\mu_2(N-j)/B_j$ 是合理的. 至于出生, 我们假设 A 型和 a 型并无不同, 因此新个体为 A 型的概率是 j/N, 新个体为 a 型的概率为 $(N-j)/N$. 考虑马尔可夫链 $\{X_n\}$, 其中 X_n 是在时刻 $t_n(n=1,2,\cdots)A$ 型个体数, 其转移概率为

$$P_{j,j-1} = \frac{\mu_1 j(N-j)}{B_j N}, \quad P_{j,j+1} = \frac{\mu_2(N-j)j}{B_j N},$$

$$P_{jj} = 1 - P_{j,j-1} - P_{j,j+1}, \quad P_{ij} = 0 \;\text{ 当 }\; |i-j| > 1\text{时}.$$

试求当初始状态为 k 个 A 型、$N-k$ 个 a 型时, 此群体最后全部变为 a 型的概率.

提示: 说明确定吸收概率的方程可以归结为赌徒输光随机游动的吸收概率所对应的方程组.

答案:

$$\begin{aligned}&\Pr\{\text{最后全部为 } a \text{ 型}\}\\&=\frac{(\mu_1/\mu_2)^N-(\mu_1/\mu_2)^k}{(\mu_1/\mu_2)^N-1},\quad \mu_1\neq\mu_2,\\&=1-\frac{k}{N},\qquad\qquad \mu_1=\mu_2.\end{aligned}$$

8. 设 $\boldsymbol{P}$ 是 3×3 马尔可夫矩阵并定义 $\mu(\boldsymbol{P})=\max\limits_{i_1,i_2,j}[P_{i_1,j}-P_{i_2,j}]$. 试证明 $\mu(\boldsymbol{P})=1$ 当且仅当 $\boldsymbol{P}$ 具有形式

$$\begin{pmatrix}1&0&0\\0&p&q\\r&s&t\end{pmatrix}\quad(p,q\geqslant0,p+q=1;r,s,t\geqslant0,r+s+t=1)$$

或由该矩阵通过调换任意两行 (列) 而得到的矩阵.

114

*9. 如果 $\boldsymbol{P}$ 是有限马尔可夫矩阵, 我们定义 $\mu(\boldsymbol{P})=\max\limits_{i_1,i_2,j}(P_{i_1,j}-P_{i_2,j})$. 设 $\boldsymbol{P}_1$, $\boldsymbol{P}_2,\cdots,\boldsymbol{P}_k$ 是 3×3 的不可约非周期马尔可夫链的转移矩阵. 进一步假设对于任意整数集合 $\alpha_i(1\leqslant\alpha_i\leqslant k),i=1,2,\cdots,m,\prod\limits_{i=1}^{m}\boldsymbol{P}_{\alpha_i}$ 也是一个非周期不可约马尔可夫链的矩阵. 试证明, 对于任给 $\varepsilon>0$, 存在一个 $M(\varepsilon)$, 使得当 $m>M$ 时, 对任何集合 $\alpha_i(1\leqslant\alpha_i\leqslant k),i=1,2,\cdots,m$, 有

$$\mu\left(\prod_{i=1}^{m}\boldsymbol{P}_{\alpha_i}\right)<\varepsilon.$$

10. 如果 i 是常返状态且 X_k 表示马尔可夫链在时刻 k 的状态, 试证明

$$\lim_{N\to\infty}\Pr\{X_k\neq i\quad\text{对于}\quad n+1\leqslant k\leqslant n+N\big|X_0=i\}=0.$$

如果 i 是正常返状态, 试证明上述收敛性关于 n 是一致的.

*11. 广义波里亚罐方案. 在一罐里有 a 个白球和 b 个黑球, 我们从中随机抽取一球, 如果抽到一个白球, 把它放回并添加 α 个白球和 β 个黑球到罐里. 如果抽到一个黑球, 也把它放回并添加 γ 个白球和 δ 个黑球到罐中. 此处, $\alpha+\beta=\gamma+\delta$. 重复此过程. 设 X_n 是前 n 次抽取中抽到白球的次数.

(i) 如果 $P_{n,k}=\Pr\{X_n=k\}$ 和 $\varphi_n(x)=\sum\limits_{k=0}^{n}P_{n,k}x^k$. 试证明恒等式

$$\varphi_n(x)=\frac{(\alpha-\gamma)(x^2-x)}{(n-1)(\alpha+\beta)+a+b}\varphi'_{n-1}(x)+\frac{\{x[(n-1)\gamma+a]+b+(n-1)\delta\}}{(n-1)(\alpha+\beta)+a+b}\varphi_{n-1}(x).$$

(ii) 证明当 $n \to \infty$ 时成立 $E(X_n/n) \to \gamma/(\beta+\gamma)$.

提示：证明

$$\varphi_n'(1) = (\alpha-\gamma)\sum_{k=1}^{n}\frac{\varphi_{k-1}'(1)}{(k-1)(\alpha+\beta)+a+b} + \sum_{k=1}^{n}\frac{a+(k-1)\gamma}{(k-1)(\alpha+\beta)+a+b}$$

并由此导出 $\varphi_n'(1)/n \to \gamma/(\beta+\gamma)$.

*12. 在问题 11 的条件下证明：当 $n\to\infty$ 时成立

$$\lim_{n\to\infty} E\left[\left(\frac{X_n}{n}\right)^2\right] = \left(\frac{\gamma}{\beta+\gamma}\right)^2.$$

115

提示：仿照上题 (i). 确定关于 $\varphi_n''(1)$ 的递推关系.

*13. 在问题 11 的条件下证明：当 $n\to\infty$ 时, X_n/n 依概率收敛于 $\gamma/(\beta+\gamma)$.

*14. 考虑一具有有限状态 $\{1,2,\cdots,N\}$ 的不可约马尔可夫链. 设 $\|P_{ij}\|$ 是此马尔可夫链的转移概率矩阵, 并用 $\{\pi_j\}$ 表示过程的平稳分布, $\|P_{ij}^{(m)}\|$ 表示 m 步转移概率矩阵. 令 $\varphi(x)$ 是 $x \geqslant 0$ 的凹函数, 定义

$$E_m = \sum_{j=1}^{N}\pi_j\varphi(P_{jl}^{(m)}), \qquad l \text{ 固定}.$$

试证明 E_m 是 m 的不减函数, 即对所有 $m\geqslant 1$ 有 $E_{m+1}\geqslant E_m$.

提示：利用 Jensen 不等式.

*15. 假设状态 0 是正常返的. 设初始状态为 0. 令 $\{W_n\}(n=1,2,\cdots)$ 表示相继的返回时间, 它们是独立同分布的随机变量, 具有有限均值和母函数 $F(t)=\sum_{k=1}^{\infty}t^k\Pr\{W_1 = k\}(|t|<1)$. 令 Y_n 表示在时刻 n 之前最后到达状态 0 的时间. 证明

$$\sum_{n=0}^{\infty}t^n\sum_{j=0}^{n}x^j\Pr\{Y_n=j\} = \frac{1-F(t)}{(1-t)(1-F(xt))}.$$

提示：证明并利用关系式

$$\Pr\{Y_n=j\} = \Pr\{W_1+\cdots+W_{N_n}=j\}.\, q_{n-j},$$

其中 $q_i=\Pr\{W_1>i\}, N_n$ 表示在时刻 n 之前到达状态 0 的次数.

16. 固定一递减非负数序列：$1=b_0\geqslant b_1\geqslant\cdots$, 考虑具有如下转移概率的马尔可夫链：

$$P_{ij} = \begin{cases} \dfrac{b_j}{b_i}(\beta_i-\beta_{i+1}), & \text{若 } j\leqslant i, \\ \dfrac{\beta_{i+1}}{\beta_i}, & \text{若 } j=i+1, \\ 0, & \text{其他}. \end{cases}$$

其中 $\beta_n = b_n/(b_1 + \cdots + b_n)$. 证明 $P_{00}^n = 1/\sigma_n$, 其中 $\sigma_n = b_1 + \cdots + b_n$. 这样, 该链为非常返的, 当且仅当 $\sum \frac{1}{\sigma_n} < \infty$.

附　　记

3.1~3.4 节论及研究马尔可夫链的重要工具, 大部分有关马尔可夫链的专著都包含这一内容.

3.5 节的例子在随机排队模型中是经典的. 详细叙述可参考 Takacs[1].

参考书目

[1] L.Takacs, *Introduction to the Theory of Queues*. Oxford Univ. Press, London and New York, 1962.

116

第 4 章 连续时间马尔可夫链的古典例子

泊松过程和生灭过程在诸如排队模型、存储模型、人口增长及工程系统等理论和应用中起着重要作用. 本章内容在每一本入门教程里均有介绍.

4.1 一般纯生过程和泊松过程

前几章研究了离散时间马尔可夫链的基本概念和方法. 在这一章里我们将对连续时间离散状态的马氏过程几个重要例子进行简要讨论.

特别地, 我们这里所处理的随机变量族 $\{X(t):0\leqslant t<\infty\}$ 中的 $X(t)$ 只能取非负整数值. 我们将限制在 $\{X(t)\}$ 是一个具有平稳转移概率的马氏过程的情形. 这样, 对任意 $t>0$, 转移概率函数

$$P_{ij}(t)=\Pr\{X(t+u)=j|X(u)=i\},\qquad i,j=0,1,2,\cdots,\tag{1.1}$$

与 $u\geqslant 0$ 无关.

在研究由某些物理现象引出的特殊随机模型过程中, 比较自然的是先确定所谓过程的无限小概率, 然后由此导出转移概率函数的显式表达式.

下面我们将由 h 较小时 $P_{ij}(h)$ 的表达式出发, 利用马尔可夫性质, 导出对任意 $t>0, P_{ij}(t)$ 所满足的微分方程组. 这个方程组的解在适当边界条件下给出 $P_{ij}(t)$的表达式. 1.2 节所介绍的泊松过程事实上正是按照这个观点来处理的.

117

为了导出一般纯生过程, 我们简单回顾一下刻划泊松过程的公理.

A. 泊松过程的公理

1.2 节已经证明了泊松过程可以由一些简单的公理所确定. 为定义相似类型的更一般过程, 我们指出泊松过程所拥有的许多进一步的性质. 特别地, 它是一个状态为非负整数的马尔可夫过程, 且具有如下性质:

(i) $\Pr\{X(t+h)-X(t)=1|X(t)=x\}=\lambda h+o(h)$ 当 $h\downarrow 0$ $(x=0,1,2,\cdots)$.

此关系式等价于

$$\lim_{h\to 0^+}\frac{\Pr\{X(t+h)-X(t)=1|X(t)=x\}}{h}=\lambda.$$

符号 $o(h)$ 意指 $\lim\limits_{h\to 0^+}o(h)/h=0$. 注意上式右边部分与 x 无关.

(ii) $\Pr\{X(t+h)-X(t)=0|X(t)=x\}=1-\lambda h+o(h)$ 当 $h\downarrow 0$.

(iii) $X(0)=0$.

这些性质由直接计算即可验证, 因为所有有关的概率表达式均是可求的.

B. 泊松过程的例子

(a) 用作说明泊松过程的一个例子是钓鱼的例子. 设随机变量 $X(t)$ 表示在时间段 $[0,t]$ 内钓到鱼的数目. 假设鱼的数量非常多, 热衷于此道者钓到鱼的机会并不比新手多, 而且某一个时刻来抢饵的鱼的数目与任何其他时刻一样多. 在这些 "理想化" 条件下, 过程 $\{X(t),t\geqslant 0\}$ 可以看作是一个泊松过程. 这个例子符合马尔可夫性 (捉到鱼的机会不依赖于已捉到鱼的数目) 和 "等候无优惠" 性, 后者是泊松过程特有的性质. 这意味着刚刚来钓鱼的人和放饵待鱼呆了四个小时而没有成功的垂钓者在往后的时间内钓到鱼的机会是均等的.

118

(b) 一个比较实际的例子是由计数理论中提出来的. 如果 $X(t)$ 是在时间区间 $[0,t]$ 内由一个 Geiger 计数器探测出的放射性原子衰变的数目. 只要该物质的半衰期比时间 t 大很多, 则该过程就是一个泊松过程. 这个规定保证了任一单位时间衰变的机会关于时间是一个常数.

(c) 泊松过程自然地出现在许多排队模型中. 在这些例子中, 人们往往更加注意那些使 $X(t)$(表示在时刻 t 的队伍长度) 产生跳跃的时刻, 而不是 $X(t)$ 本身的取值. 当然, 钓鱼的例子 (a) 就是一个等候时间的典型例子.

C. 纯生过程

泊松过程的一个自然推广是允许一个事件在给定时刻发生的机会依赖于已发生的事件数目. 这种现象的一个例子是生物群体的增殖 (纯生过程的名称即源于此), 在一定的条件下, 例如有足够的食物, 没有大量死亡, 没有向外迁移等, 给定时刻增殖一个新成员的概率与群体当时的数量成正比. 这个例子就是著名的 Yule 过程.

考虑一列正数 $\{\lambda_k\}$, 我们定义纯生过程为满足如下条件的马氏过程:

(i) $\Pr\{X(t+h)-X(t)=1|X(t)=k\}=\lambda_k h+o_{1,k}(h),\ (h\to 0^+)$,

(ii) $\Pr\{X(t+h)-X(t)=0|X(t)=k\}=1-\lambda_k h+o_{2,k}(h)$,

(iii) $\Pr\{X(t+h)-X(t)<0|X(t)=k\}=0,\ (k\geqslant 0)$.

为了方便, 通常附加假设

(iv) $X(0)=0$.

在这个假定下, $X(t)$ 不表示群体总量, 而是表示在时间区间 $[0,t]$ 内群体增殖数目.

注意 (i) 和 (ii) 的左边部分刚好分别是 $P_{k,k+1}(h)$ 和 $P_{k,k}(h)$(由于平稳性), 可以 $o_{1,k}(h)$ 和 $o_{2,k}(h)$ 不依赖于 t.

设 $X(0)=0$ 并令 $P_n(t)=\Pr\{X(t)=n\}$.

119

与泊松过程情况相同, 可以推导出当 $t \geqslant 0$ 时 $P_n(t)$ 所满足的微分方程组, 即

$$\begin{aligned} &P_0'(t) = -\lambda_0 P_0(t), \\ &P_n'(t) = -\lambda_n P_n(t) + \lambda_{n-1} P_{n-1}(t), \quad n \geqslant 1, \end{aligned} \tag{1.2}$$

并伴有边界条件

$$P_0(0) = 1 \quad 和 \quad P_n(0) = 0, \quad 当\ n > 0.$$

事实上, 如果 $h > 0$, $n \geqslant 1$, 则利用全概率公式、马氏性和假设 (iii), 我们得

$$\begin{aligned} P_n(t+h) &= \sum_{k=0}^{\infty} P_k(t) \Pr\{X(t+h) = n | X(t) = k\} \\ &= \sum_{k=0}^{\infty} P_k(t) \Pr\{X(t+h) - X(t) = n-k | X(t) = k\} \\ &= \sum_{k=0}^{n} P_k(t) \Pr\{X(t+h) - X(t) = n-k | X(t) = k\}. \end{aligned}$$

对于 $k = 0, 1, \cdots, n-2$, 我们有

$$\begin{aligned} &\Pr\{X(t+h) - X(t) = n-k | X(t) = k\} \\ &\quad \leqslant \Pr\{X(t+h) - X(t) \geqslant 2 | X(t) = k\} \\ &\quad = o_{1,k}(h) + o_{2,k}(h) \end{aligned}$$

或

$$\Pr\{X(t+h) - X(t) = n-k | X(t) = k\} = o_{3,n,k}(h), \quad k = 0, \cdots, n-2.$$

从而

$$\begin{aligned} P_n(t+h) = &P_n(t)[1 - \lambda_n h + o_{2,n}(h)] \\ &+ P_{n-1}(t)[\lambda_{n-1} h + o_{1,n-1}(h)] \\ &+ \sum_{k=0}^{n-2} P_k(t) o_{3,n,k}(h) \end{aligned}$$

或

$$\begin{aligned} P_n(t+h) - P_n(t) = &P_n(t)[-\lambda_n h + o_{2,n}(h)] \\ &+ P_{n-1}(t)[\lambda_{n-1} h + o_{1,n-1}(h)] + o_n(h), \end{aligned} \tag{1.3}$$

其中 $o_n(h)$ 不大于有限和 $\sum_{k=0}^{n-2} o_{3,n,k}(h)$, 后者与 t 无关, 因此极限 $\lim_{h \downarrow 0} o_n(h)/h = 0$ 关于 $t \geqslant 0$ 一致成立.

用 h 除 (1.3) 式两端, 并令 $h\downarrow 0$, 我们导出关系式 (1.2). 准确地说, 其中左式是右导数. 然而, 稍加留意我们可以导出对于左导数也有相同关系式. 事实上, 从 (1.3) 我们立刻可知 $P_n(t)$ 是 t 的连续函数. 在 (1.3) 中用 $t-h$ 代替 t, 然后用 h 除之, 并令 $h\downarrow 0$, 我们即可得到 $P_n(t)$ 存在左导数并且满足方程 (1.2).

120

由 (1.2) 的第一个方程立刻可求解得

$$P_0(t)=\exp(-\lambda_0 t)>0.$$

以 T_k 表示第 k 次和第 $k+1$ 次增殖之间的间隔时间, 则

$$P_n(t)=\Pr\Big\{\sum_{i=0}^{n-1}T_i\leqslant t<\sum_{i=0}^{n}T_i\Big\}.$$

随机变量 T_k 称为两次增殖之间的 "等候时间", 并且

$$S_k=\sum_{i=0}^{k-1}T_i=\text{ 第 } k \text{ 次增殖发生的时间}.$$

我们已见到 $P_0(t)=\exp(-\lambda_0 t)$. 因此,

$$\Pr\{T_0\leqslant z\}=1-\Pr\{X(z)=0\}=1-\exp(-\lambda_0 z),$$

即, T_0 具有以 λ_0 为参数的指数分布. 可以由假设 (i)~(iv) 推得 $T_k(k>0)$ 也具有以 λ_k 为参数的指数分布, 并且 T_i 是相互独立的 (见第 II 卷第 14 章, 在那里将给出这个事实的严格证明). 所以, S_n 的特征函数为

$$\varphi_n(w)=E\{\exp(\mathrm{i}wS_n)\}=\prod_{k=0}^{n-1}E(\exp(\mathrm{i}wT_k))=\prod_{k=0}^{n-1}\frac{\lambda_k}{\lambda_k-\mathrm{i}w}. \tag{1.4}$$

在泊松过程情形下, 对所有的 $k,\lambda_k=\lambda$, 由 (1.4) 我们可知这时 S_n 是服从均值为 n/λ 的 n 阶 Γ 分布.

对于指定的一列 $\lambda_k\geqslant 0$, 我们可以使用积分因子 $\exp(\lambda_k t)$ 解 (1.2), 得到

$$P_k(t)=\lambda_{k-1}\exp(-\lambda_k t)\int_0^t\exp(\lambda_k x)P_{k-1}(x)\mathrm{d}x,\quad k=1,2,\cdots,$$

显然, 对所有的 k 有 $P_k(t)\geqslant 0$.

但还有可能使得

$$\sum_{n=0}^{\infty}P_n(t)<1.$$

为保证过程是合理的, 尚要确定使 $\sum_{n=0}^{\infty} P_n(t) = 1$ 对所有 $t \geqslant 0$ 成立的条件. 我们必须按照下面关系式限制 λ_k:

121

$$\sum_{n=0}^{\infty} P_n(t) = 1 \quad \Longleftrightarrow \sum_{n=0}^{\infty} \frac{1}{\lambda_n} = \infty. \tag{1.5}$$

这个结论在 Feller 的书中已证明 [1] , 这里从略. 关于此结果直观的论证可叙述如下: 相邻的两次增殖之间的时间 T_k 已证明是具有参数为 λ_k 的指数分布, 所以 $\sum_n 1/\lambda_n$, 等于群体总量变为无限之前的期望时间. 另一方面, $1 - \sum_{n=0}^{\infty} P_n(t)$ 是 $X(t) = \infty$ 的概率.

假若 $\sum \lambda_n^{-1} < \infty$, 则群体总量变为无限的期望时间是有限的, 因而对所有 $t > 0$, $X(t) = \infty$ 的概率为正的似乎有理.

D. Yule 过程

Yule 过程是纯生过程的一个例子, 它是从物理学和生物学中提出来的. 假设在一个群体中每个成员在长度为 h 的时间区间内增殖一个新的成员的概率为 $\beta h + o(h)(\beta > 0)$. 此外, 设在时刻 $t = 0$ 存在 $X(0) = N$ 个成员, 并假定群体成员之间相互独立, 不存在相互作用. 由二项定理得

$$\begin{aligned} &\Pr\{X(t+h) - X(t) = 1 | X(t) = n\} \\ &= \binom{n}{1} [\beta h + o(h)][1 - \beta h + o(h)]^{n-1} = n\beta h + o_n(h), \end{aligned}$$

即, 在这个例子中 $\lambda_n = n\beta$. 方程组 (1.2) 当 $N = 1$ 时成为

$$P_n'(t) = -\beta[nP_n(t) - (n-1)P_{n-1}(t)], \qquad n = 1, 2, \cdots,$$

其初始条件为

$$P_1(0) = 1, \ P_n(0) = 0, \ n = 2, 3, \cdots,$$

它的解可直接验证是

$$P_n(t) = \mathrm{e}^{-\beta t}(1 - \mathrm{e}^{-\beta t})^{n-1}, n \geqslant 1,$$

---

1. W.Feller, *An Introduction to Probability Theory and Its Applications*, Vol.1, 2nd ed. p.406. Wiley, New York, 1957.

通过求几何级数之和, 易得其母函数为

$$
\begin{aligned}
f(s) &= \sum_{n=1}^{\infty} P_n(t) s^n \\
&= s\mathrm{e}^{-\beta t} \sum_{n=1}^{\infty} [(1-\mathrm{e}^{-\beta t})s]^{n-1} = \frac{s\mathrm{e}^{-\beta t}}{1-(1-\mathrm{e}^{-\beta t})s}.
\end{aligned}
$$

122

现在我们回到一般情况, 设在时刻 0 有成员 $X(0) = N$ 个. 由于假定各成员彼此独立而没有相互作用, 我们可以把这个总数看为 N 个独立 Yule 过程的和, 其中每一个过程在开始时刻只有单独一个成员. 如果我们令

$$
P_{Nn}(t) = \Pr\{X(t) = n | X(0) = N\}
$$

和

$$
f_N(s) = \sum_{n=N}^{\infty} P_{Nn}(t) s^n, \tag{1.6}
$$

使得

$$
\begin{aligned}
f_N(s) = [f(s)]^N &= \left[\frac{s\mathrm{e}^{-\beta t}}{1-(1-\mathrm{e}^{-\beta t})s}\right]^N \\
&= (s\mathrm{e}^{-\beta t})^N \sum_{m=0}^{\infty} \binom{m+N-1}{m} (1-\mathrm{e}^{-\beta t})^m s^m \\
&= \sum_{n=N}^{\infty} \binom{n-1}{n-N} (\mathrm{e}^{-\beta t})^N (1-\mathrm{e}^{-\beta t})^{n-N} s^n,
\end{aligned}
$$

这里, 我们利用了二项级数 $(1-x)^{-N} = \sum_{m=0}^{\infty} \binom{m+N-1}{m} x^m$. 按照 (1.6), 在这个表达式中 s^n 的系数必须是 $P_{Nn}(t)$, 即

$$
P_{Nn}(t) = \binom{n-1}{n-N} \mathrm{e}^{-N\beta t} (1-\mathrm{e}^{-\beta t})^{n-N}, \quad n = N, N+1, \cdots. \tag{1.7}
$$

4.2 泊松过程的补充

在前一节我们从一组在许多实际情况中都能很好逼近的假设出发导出了泊松过程. 它的分布点如同有限区间上的均匀分布那样 "随机地" 落在无限区间 $[0,\infty)$ 上, 此过程堪称为完全的随机过程. 特别地, 一个观测值落在某个子区间的概率仅仅是此区间长度的函数, 并且在不相交时间区间里事件出现的次数是相互独立的随机变量.

下面我们更细微地研究泊松过程.

123

A. 特征函数和等候时间

泊松过程中 $X(t)$ 的特征函数为

$$\varphi_t(w) = E\{e^{iwX(t)}\} = \sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n e^{iwn}}{n!} = \exp[\lambda t(e^{iw} - 1)]$$

于是,

$$E(X(t)) = \lambda t, \quad \mathrm{Var}(X(t)) = \lambda t.$$

在纯生过程的讨论中, 我们证明了

$$\Pr\{T_0 \leqslant z\} = 1 - \exp(-\lambda_0 z),$$

并且指出了 T_k 遵循参数为 λ_k 的指数分布, $T_k(k \geqslant 0)$ 是相互独立的. 然而, 泊松过程对所有 k 成立 $\lambda_k = \lambda$, 故有

定理 2.1 对于泊松过程, 等候时间 T_k 是独立同分布的, 且服从参数为 λ 的指数分布.

这个定理的严格证明将由第 14 章更一般的结果导出.

泊松过程的定义比此定理成立的条件要求更多一些. 我们需要假定从任何时刻开始至 $X(t)$ 下一次改变的时间服从相同的分布, 而不必一定以上次改变为初始时刻, 用数学式子表达是

$$\Pr\{X(t_0 + t) - X(t_0) > 0\} = 1 - e^{-\lambda t},$$

此式已在 4.1 节导出, 这个性质也可以用更加直接的方法得到. 令 $F(x) = \Pr\{X(t_0 + x) - X(t_0) > 0\}$, 其中 t_0 是某个时刻, 也可以是一随机时间仅依赖于至此时刻为止过程的历史, 它的具体选择并不影响此概率[1]. 我们有

$$\begin{aligned} F(x+y) &= \Pr\{X(t_0 + x + y) - X(t_0) > 0\} \\ &= \Pr\{X(t_0 + y) - X(t_0) > 0\} + \Pr\{X(t_0 + y) - X(t_0) = 0\} \\ &\quad \times \Pr\{X(t_0 + x + y) - X(t_0 + y) > 0 | X(t_0 + y) - X(t_0) = 0\}. \end{aligned}$$

124 由 $F(x)$ 的定义, 泊松过程增量的独立性以及 $\Pr\{X(t_0 + x) - X(t_0) > 0\}$ 与 t_0 无关 (在泊松过程定义中作为一个假设), 我们得到函数方程

$$F(x+y) = F(y) + [1 - F(y)]F(x).$$

事实上, 这个性质所描述的指数分布的特征正是如下定理的内容:

1. 这里显得含糊不清, 我们将在第 14 章讨论 “马尔可夫时间” 概念时给予比较准确的解释.

定理 2.2 如果 $F(x)$ 是一个分布, $F(0)=0$, 且对于某个 $x>0$ 成立 $F(x)<1$, 则 $F(x)$ 为指数分布的充分必要条件是对于所有的 $x,y\geqslant 0$, 成立

$$F(x+y)-F(y)=F(x)[1-F(y)]. \tag{$*$}$$

证明 指数分布应满足的这个条件可直接验证. 为说明充分性, 置 $G(x)=1-F(x)$; 则条件 $(*)$ 变为

$$G(x+y)=G(x)G(y) \tag{2.1}$$

显然, $G(0)=1$, $G(x)$ 是不增的, 并且对某个 $x>0$, $G(x)>0$ 假设对某个 $x_0>0$, $G(x_0)=0$. 由方程 (2.1) 推得 $G(x_0)=[G(x_0/n)]^n$ 对每个正整数 n 成立; 因此 $G(x_0/n)=0$. 故由 (2.1) 知对于 $x>x_0/n$ 有 $G(x)=0$. 既然 n 是任意的, 因而对所有 $x>0$ 有 $G(x)=0$, 这与假设矛盾. 从而对每个 $x>0$ 成立 $G(x)>0$. 现对任意正整数 m,n, 从 (2.1) 容易推出 $G(m/n)=[G(1)]^{m/n}$. 由于 $G(x)$ 和 $[G(1)]^x$ 皆是不增函数, 当 x 为有理数时二者相等, 且 $[G(1)]^x$ 是连续的, 我们推得对所有 $x>0$ 成立 $G(x)=[G(1)]^x=\exp(x\ln G(1))$. 但 $F(x)$ 是一个分布, 故

$$\lim_{x\to\infty}G(x)=1-\lim_{x\to\infty}F(x)=0,$$

此式蕴含着 $G(1)<1$, 因此 $G(x)=\mathrm{e}^{-\lambda x}$, 其中

$$\lambda=-\ln G(1)>0.$$

■

当假定 G 是可微分时, 另一种方法证明如下: 显然 (2.1) 蕴涵

$$G'(x+y)=\frac{\partial}{\partial x}G(x+y)=G'(x)G(y),$$
$$G(x)G'(y)=\frac{\partial}{\partial y}G(x+y)=G'(x+y).$$

125

故

$$G'(x)=aG(x), \tag{2.2}$$

其中 $a=G'(y_0)/G(y_0)$, y_0 为某个使 $G(y_0)\neq 0$ 的常值. 方程 (2.2) 的解为 $G(x)=A\mathrm{e}^{ax}$. 由 $G(0)=1-F(0)=1$ 可知 $A=1$. 由于对某个 $x>0$ 有 $G(x)<1$, 因此参数 a 是负的.

B. 均匀分布

与泊松过程相联系的分布不仅有泊松分布和指数分布, 还有均匀分布和二项分布.

考虑一列时刻 $\{S_i\}$, 在每个时刻 S_i, $X(t)$ 值发生变化, 即

$$S_i = \sum_{k=0}^{i-1} T_k.$$

我们有下面结果

定理 2.3 对满足 $0 \leqslant s_1 \leqslant s_2 \leqslant \cdots \leqslant s_n \leqslant t$ 的任何一组数 s_i, 有

$$\begin{aligned}&\Pr\{S_i \leqslant s_i, i = 1, 2, \cdots, n | X(t) = n\}\\&= \frac{n!}{t^n} \int_0^{s_1} \cdots \int_{x_{n-2}}^{s_{n-1}} \int_{x_{n-1}}^{s_n} \mathrm{d}x_n \cdots \mathrm{d}x_1,\end{aligned}$$

这是从具有 $[0, t]$ 上均匀分布的总体取 n 个观测样本的顺序统计量分布[1].

证明 由定理 2.1 容易推出本定理结论. 事实上,

$$\begin{aligned}&\Pr\{S_1 \leqslant s_1, S_2 \leqslant s_2, \cdots, S_n \leqslant s_n, X(t) = n\}\\&= \Pr\{T_0 \leqslant s_1, T_0 + T_1 \leqslant s_2, \cdots, T_0 + \cdots + T_{n-1} \leqslant s_n, T_0 + \cdots + T_n > t\}\\&= \int_0^{s_1} \int_0^{s_2 - t_1} \int_0^{s_3 - (t_1 + t_2)} \cdots \int_0^{s_n - (t_1 + \cdots + t_{n-1})} \int_{t - (t_1 + \cdots + t_n)}^{\infty}\\&\quad \lambda^{n+1} \mathrm{e}^{-\lambda(t_1 + \cdots + t_{n+1})} \mathrm{d}t_{n+1} \cdots \mathrm{d}t_1\\&= \lambda^{n+1} \int_0^{s_1} \int_0^{s_2 - t_1} \int_0^{s_3 - (t_1 + t_2)} \cdots \int_0^{s_n - (t_1 + \cdots + t_{n-1})}\\&\quad \mathrm{e}^{-\lambda(t_1 + \cdots + t_n)} \left[-\frac{1}{\lambda} \exp(-\lambda t_{n+1}) \right]_{t - (t_1 + \cdots + t_n)}^{\infty} \mathrm{d}t_n \cdots \mathrm{d}t_1\\&= \lambda^n \mathrm{e}^{-\lambda t} \int_0^{s_1} \int_0^{s_2 - t_1} \int_0^{s_3 - (t_1 + t_2)} \cdots \int_0^{s_n - (t_1 + \cdots + t_{n-1})} \mathrm{d}t_n \cdots \mathrm{d}t_1.\end{aligned}$$

126

若我们引进新的变量

$$\begin{aligned}u_n &= t_1 + \cdots + t_n\\u_{n-1} &= t_1 + \cdots + t_{n-1}\\&\vdots\\u_1 &= t_1,\end{aligned}$$

最后的表达式变为

$$\lambda^n \mathrm{e}^{-\lambda t} \int_0^{s_1} \int_{u_1}^{s_2} \int_{u_2}^{s_3} \cdots \int_{u_{n-1}}^{s_n} \mathrm{d}u_n \cdots \mathrm{d}u_1.$$

1. 这意味着对一个有 $[0, t]$ 上均匀分布的随机变量进行 n 次独立观察, 令 $Y_1 \leqslant Y_2 \leqslant \cdots \leqslant Y_n$ 表示这些观察值按照增加顺序安排. 那么, $Y_1, Y_2, \cdots, Y_n$ 的联合分布正是定理结论中的表达式. 这个事实证明是十分简单的, 更加复杂的讨论见第 II 卷第 13 章.

但

$$\Pr\{X(t)=n\}=\mathrm{e}^{-\lambda t}\frac{(\lambda t)^n}{n!},$$

因此,

$$\begin{aligned}&\Pr\{S_1\leqslant s_1,S_2\leqslant s_2,\cdots,S_n\leqslant s_n|X(t)=n\}\\=&\frac{\Pr\{S_1\leqslant s_1,\cdots,S_n\leqslant s_n,X(t)=n\}}{\Pr\{X(t)=n\}}\\=&\frac{n!}{t^n}\int_0^{s_1}\int_{u_1}^{s_2}\cdots\int_{u_{n-1}}^{s_n}\mathrm{d}u_n\cdots\mathrm{d}u_1.\end{aligned}$$

C. 二项分布

由泊松过程的性质知, 若 $u<t$ 和 $k<n$,

$$\Pr\{X(u)=k|X(t)=n\}=\Pr\{X(u)=k,X(t)-X(u)=n-k\}/\Pr\{X(t)=n\}$$

$$=\frac{(\mathrm{e}^{-\lambda u}u^k/k!)[\mathrm{e}^{-\lambda(t-u)}(t-u)^{n-k}/(n-k)!]}{\mathrm{e}^{-\lambda t}(t^n/n!)}=\binom{n}{k}\frac{u^k(t-u)^{n-k}}{t^n}\qquad(2.3)$$

127

与二项分布有关的第二个例子可以考虑两个独立泊松过程 $X_1(t)$ 和 $X_2(t)$, 其参数分别是 λ_1 和 λ_2,

$$\begin{aligned}&\Pr\{X_1(t)=k|X_1(t)+X_2(t)=n\}=\frac{\Pr\{X_1(t)=k,X_2(t)=n-k\}}{\Pr\{X_1(t)+X_2(t)=n\}}\\=&\frac{[\exp(-\lambda_1 t)(\lambda_1 t)^k/k!][\exp(-\lambda_2 t)(\lambda_2 t)^{n-k}/(n-k)!]}{\exp[-(\lambda_1+\lambda_2)t](\lambda_1+\lambda_2)^n t^n/n!}\\=&\binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}.\end{aligned}$$

4.3 计数模型

下面的问题是泊松过程的一个有趣的应用: 具有随机振幅为 X_i 的电子脉冲在随机时间 t_i(即按照泊松过程) 到达一个探测器. 对于每一个脉冲在时刻 t 探测器的输出是:

$$X_i\exp[-\alpha(t-t_i)]_+=\begin{cases}0 & t<t_i,\\ X_i\exp[-\alpha(t-t_i)] & t\geqslant t_i;\end{cases}$$

即当脉冲到达时, 以振幅 X_i 传送到探测器, 而其输出是以指数率衰减. 由于探测器是线性的 (即可加的), 所以若在时间区间 $[0,t]$ 有 N_t 个脉冲出现, 则在时间 t 的输出是

$$\eta(t)=\sum_{i=1}^{N_t}X_i\exp[-\alpha(t-t_i)]_+.$$

图 4-1 给出这个过程的一个具体实现, 我们希望能知道 $\eta(t)$ 对每个 t 的分布函数, 或等价地, 知道它的特征函数 $\varphi_t(w)$.

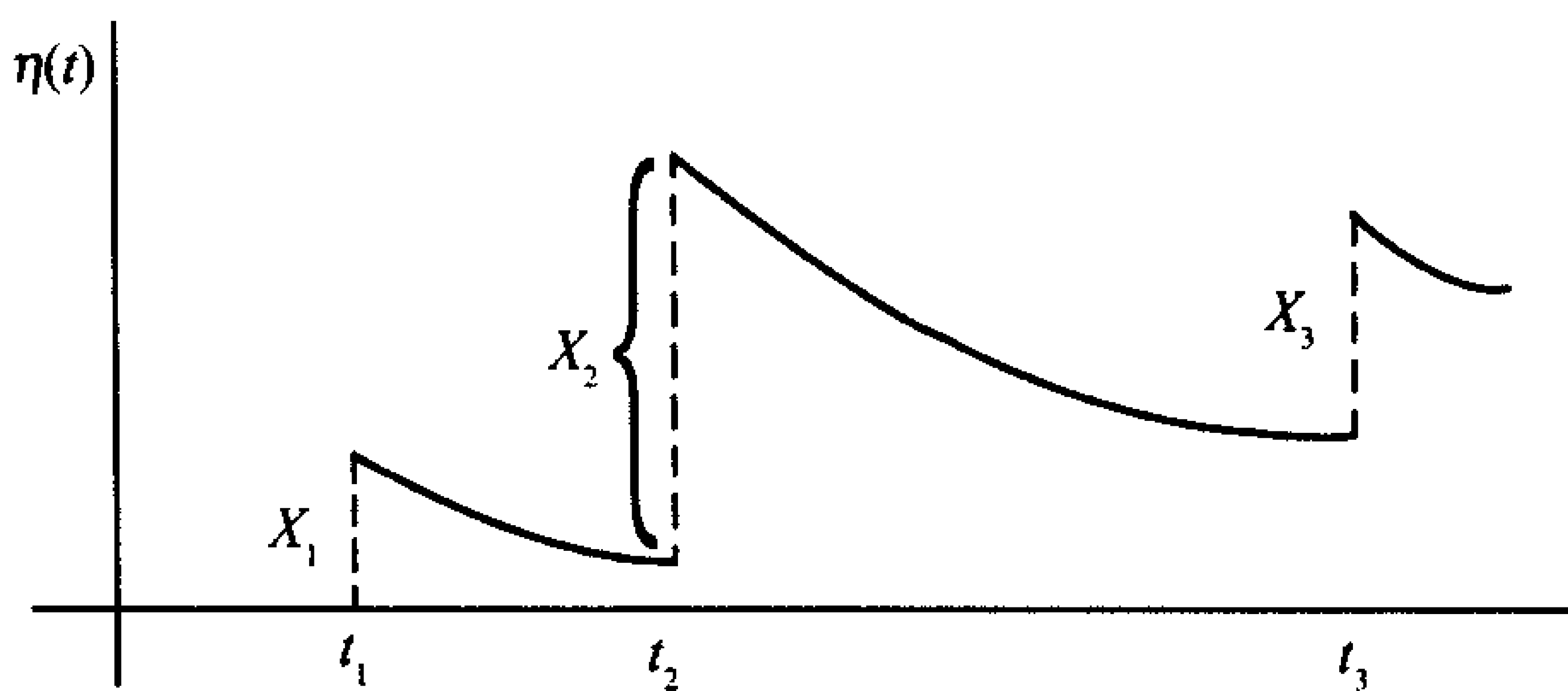


图　4-1

假设 X_i 是独立同分布且具有密度函数 $h(x)$ 的正随机变量, 其特征函数为

$$\psi(s) = \int_0^\infty e^{isx}h(x)dx.$$

置

128

$$R(v;t) = \Pr\{\eta(t) \leqslant v\} = \sum_{n=0}^{\infty} \Pr\{\eta(t) \leqslant v | N_t = n\}\Pr\{N_t = n\}. \tag{3.1}$$

显然 $\Pr\{N_t = n\} = [(\lambda t)^n e^{-\lambda t}]/n!$, 其中 λ 是描述脉冲到达时间的泊松过程强度参数. 由定理 2.3 知, 在条件 $N_t = n$, 即时间区间 $(0,t)$ 内有 n 个脉冲到达的情况下, t_i 好像是对 $(0,t)$ 上均匀分布进行 n 次观测所得到顺序值. 令 $\tau_i(i = 1,2,\cdots,N_t)$ 表示具有 $(0,t)$ 上均匀分布的独立随机变量, 它若按大小顺序安排, 其值即为 $t_j(j = 1,2,\cdots,n)$.

现令 $Z_1,\cdots,Z_n$ 是独立随机变量, 其分布与 X_i 是相同的, 并且与 $\{\tau_i\}$ 是独立的. 考虑和

$$\sum_{i=1}^{n} Z_i \exp[-\alpha(t-\tau_i)]_+.$$

我们定义新的随机变量 $Z'_1,\cdots,Z'_n$ 如下

$$\begin{aligned}
&Z'_1 = Z_j \quad 当\ \tau_j = \min(\tau_1,\cdots,\tau_n) = t_1\\
&Z'_2 = Z_j \quad 当\ \tau_j\ 是在\ \{\tau_j\}\ 中除\ t_1\ 之外最小\ = t_2\\
&\ \vdots\ \ \ \vdots \qquad \vdots\\
&Z'_n = Z_j \quad 当\ \tau_j = \max(\tau_1,\cdots,\tau_n) = t_n
\end{aligned}$$

当两个或更多个 τ_i 相等时, 上面定义不明确, 此时无关紧要, 因为这些事件发生的概率为 0. 那么,

$$\sum_{i=1}^{n} Z_i \exp[-\alpha(t-\tau_i)]_+ = \sum_{i=1}^{n} Z_i' \exp[-\alpha(t-t_i)]_+,$$

因为这两个和仅差一个随机重排. 因为 Z_i 是独立同分布的, 且独立于 τ_i, 则容易证明 Z_i' 是独立的, 它们的分布与 Z_i 分布相同, 且也与 τ_i 无关. 由于独立于 τ_i, $\{Z_i\}$ 和 $\{Z_i'\}$ 显然也独立于 t_i. 129

因为 Z_i' 具有 X_i 所需要的全部性质, 所以我们可以取

$$\eta(t) = \sum_{i=1}^{n} Z_i' \exp[-\alpha(t-t_i)]_+ = \sum_{i=1}^{n} Z_i \exp[-\alpha(t-\tau_i)]_+.$$

令

$$Y_t(i) = Z_i \exp[-\alpha(t-t_i)]_+;$$

显然对固定 t, $Y_t(i), i=1,2,\cdots,n$ 是独立同分布随机变量. 定义

$$\theta_t(s) = \int_0^{\infty} \mathrm{e}^{\mathrm{i}sy} g_t(y,k)\mathrm{d}y,$$

为 $Y_t(k)$ 的特征函数, 其中 $g_t(y,k)$ 是 $Y_t(k)$ 的密度函数. 由于 τ_k 是 $(0,t)$ 上均匀分布, 且 τ_k 和 Z_k 是独立的, 我们有

$$\begin{aligned}
\int_0^y g_t(u,k)\mathrm{d}u &= \Pr\{Y_t(k) \leqslant y\} \\
&= \Pr\{Z_k \exp[-\alpha(t-\tau_k)]_+ \leqslant y\} \\
&= \int_0^t \Pr\{Z_k \exp[-\alpha(t-\tau_k)]_+ \leqslant y | \tau_k = u\} \frac{\mathrm{d}u}{t} \\
&= \int_0^t \Pr\{Z_k \leqslant y\mathrm{e}^{\alpha(t-u)}\} \frac{\mathrm{d}u}{t} \\
&= \frac{1}{t}\int_0^t H(y\mathrm{e}^{\alpha(t-u)})\mathrm{d}u,
\end{aligned} \tag{3.2}$$

其中 H 是对应于密度 h 的分布函数. 微分 (3.2) 得到

$$g_t(y,k) = \frac{1}{t}\int_0^t h(y\mathrm{e}^{\alpha(t-u)})\mathrm{e}^{\alpha(t-u)}\mathrm{d}u.$$

所以

$$\theta_t(s) = \int_0^\infty e^{isy} g_t(y,k)dy = \frac{1}{t}\int_0^t e^{\alpha(t-u)}\left(\int_0^\infty e^{isy} h(ye^{\alpha(t-u)})dy\right)du$$

130

$$= \frac{1}{t}\int_0^t du \int_0^\infty \exp[is(e^{-\alpha(t-u)}z)]h(z)dz \qquad (\text{进行变量替换 } ye^{\alpha(t-u)} = z)$$

$$= \frac{1}{t}\int_0^t \psi(se^{-\alpha t}e^{\alpha u})du \qquad (\text{由 } \psi \text{ 的定义})$$

$$= \frac{1}{t}\int_0^t \psi(se^{-\alpha v})dv \qquad (v = t-u)$$

由此推得, 如果 $r(x,t)$ 是 $R(x,t)$ 的密度函数, 则

$$\varphi_t(w) = \int_0^\infty e^{iwx} r(x,t)dx$$

$$= \sum_{n=0}^\infty \left(\int_0^\infty e^{iwx}\frac{d}{dx}\Pr\{\eta(t) \leqslant x | N_t = n\}dx\right)e^{-\lambda t}\frac{(\lambda t)^n}{n!} \quad [\text{利用 } (3.1)]$$

$$= \sum_{n=0}^\infty e^{-\lambda t}\frac{(\lambda t)^n}{n!}[\theta_t(w)]^n \qquad (\text{利用在已知 } N_t \text{ 值条件下 } Y_t(k) \text{ 的独立性})$$

$$= \sum_{n=0}^\infty \frac{e^{-\lambda t}}{n!}\left(\lambda \int_0^t \psi(we^{-\alpha v})dv\right)^n$$

$$= \exp - \left\{\lambda \int_0^t [1 - \psi(we^{-\alpha v})]dv\right\}.$$

关于 w 微分, 我们可计算 $\eta(t)$ 的矩, 例如

$$E(\eta(t)) = (-i)\frac{d}{dw}\varphi_t(w)|_{w=0} = \lambda(-i)\psi'(0)\int_0^t e^{-\alpha v}dv$$

$$= \lambda \cdot E(X_k)\frac{1-e^{-\alpha t}}{\alpha}.$$

4.4　生灭过程

4.1 节所讨论纯生过程的一个自然推广是 $X(t)$ 除可增加之外还可减少, 例如成员的死亡. 这样, 如果在时刻 t 过程处于状态 n, 经过一个随机等候时间之后, 它可以转移到相邻的状态 $n+1$ 或 $n-1$ 中之一. 因此 "生灭过程" 可以比拟为连续时间情况下的随机游动 (2.2 节例 B).

131

A. 假设

如同纯生过程的情况, 我们假定 $X(t)$ 是状态为非负整数的马尔可夫过程, 且它的转移概率是平稳的, 即

$$P_{ij}(t) = \Pr\{X(t+s) = j | X(s) = i\}.$$

不依赖于 s. 此外, 我们假定 $P_{ij}(t)$ 满足

1. $P_{i,i+1}(h)=\lambda_i h+o(h)$ 当 $h\downarrow 0, i\geqslant 0$;
2. $P_{i,i-1}(h)=\mu_i h+o(h)$ 当 $h\downarrow 0, i\geqslant 1$;
3. $P_{i,i}(h)=1-(\lambda_i+\mu_i)h+o(h)$ 当 $h\downarrow 0,\ i\geqslant 0$;
4. $P_{ij}(0)=\delta_{ij}$;
5. $\mu_0=0,\lambda_0>0,\mu_i,\lambda_i>0,i=1,2,\cdots$.

在上面每种情况下, $o(h)$ 可能与 i 有关. 矩阵

$$\boldsymbol{A}=\left\|\begin{matrix}-\lambda_0 & \lambda_0 & 0 & 0 & \cdots\\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & 0 & \cdots\\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & \cdots\\ 0 & 0 & \mu_3 & -(\lambda_3+\mu_3) & \cdots\\ \vdots & \vdots & \vdots & \vdots & \end{matrix}\right\| \tag{4.1}$$

称为过程的**无穷小生成元**. 参数 λ_i 和 μ_i 分别被称为无穷小生 (或增长) 和灭 (或消亡) 的比率. 在假设 1 和 2 中, 我们假定了过程初始处于状态 i, 则在一个小的时间区间内, 群体增加和减少实际上与区间的长度成比例. 有时也允许状态零为吸收状态 (见 4.7 节).

由于 $P_{ij}(t)$ 是概率, 我们有 $P_{ij}(t)\geqslant 0$ 和

$$\sum_{j=0}^{\infty}P_{ij}(t)=1 \tag{4.2}$$

利用过程的马氏性还可导出切普曼 — 科尔莫戈罗夫方程

$$P_{ij}(t+s)=\sum_{k=0}^{\infty}P_{ik}(t)P_{kj}(s) \tag{4.3}$$

这个方程表示为了从状态 i 经过时间 $t+s$ 转移到状态 j, $X(t)$ 可经过时间 t 先转移到某个状态 k, 然后在余下时间 s 再从 k 转移到 j. 在连续时间情况的这个公式类似于第 2 章公式 (3.2).

132

迄今, 我们仅提及转移概率 $P_{ij}(t)$. 为了得到 $X(t)=n$ 的概率, 尚必须规定过程从什么状态出发, 或更一般地, 关于初始状态的概率分布. 然后, 我们才有

$$\Pr\{X(t)=n\}=\sum_{i=0}^{\infty}q_iP_{in}(t),$$

其中

$$q_i=\Pr\{X(0)=i\}.$$

B. 等候时间

借助于上述假设, 我们来求随机变量 T_i 的分布, T_i 是过程 $X(t)$ 在状态 i 的等候时间. 这里的问题就是, 已知过程处于状态 i, 如何求首次离开 i 的时间 T_i 的分布? 令

$$\Pr(T_i \geqslant t) = G_i(t)$$

由马尔可夫性易推得当 $h \downarrow 0$ 时有

$$G_i(t+h) = G_i(t)G_i(h) = G_i(t)(P_{ii}(h) + o(h)) = G_i(t)[1 - (\lambda_i + \mu_i)h] + o(h)$$

或

$$\frac{G_i(t+h) - G_i(t)}{h} = -(\lambda_i + \mu_i)G_i(t) + o(1),$$

所以

$$G_i'(t) = -(\lambda_i + \mu_i)G_i(t). \tag{4.4}$$

如果我们利用条件 $G_i(0) = 1$, 这个方程的解是

$$G_i(t) = \exp[-(\lambda_i + \mu_i)t],$$

即 T_i 服从均值为 $(\lambda_i + \mu_i)^{-1}$ 的指数分布. 上述提供的证明不是十分严格的, 因为我们未加证明地利用了直观关系式

$$G_i(h) = P_{ii}(h) + o(h).$$

关系式 (4.4) 的严格证明将在第 2 卷第 14 章给出.

按照假设 1 和 2, 在长度为 h 的时间区间内, 发生一次转移使过程从状态 i 到达状态 $i+1$ 的概率为 $\lambda_i h + o(h)$, 从状态 i 到达状态 $i-1$ 的概率为 $\mu_i h + o(h)$. 由直观可知, 若已知转移发生在时刻 t, 则转移到状态 $i+1$ 的概率是 $\lambda_i(\mu_i+\lambda_i)^{-1}$, 转移到状态 $i-1$ 的概率是 $\mu_i(\mu_i+\lambda_i)^{-1}$. 这个结果的严格证明已超出本书范围, 然而关于此问题的解释及其内在精细之处将在后面给出 (第 2 卷第 14 章).

[133]

$X(t)$ 的运动可如下描述: 过程在给定状态 i 逗留一段时间, 其长度是服从参数为 $\lambda_i + \mu_i$ 的指数分布的随机变量, 然后离开状态 i, 并分别以概率 $\lambda_i(\mu_i+\lambda_i)^{-1}$ 和 $\mu_i(\mu_i+\lambda_i)^{-1}$ 到达状态 $i+1$ 或状态 $i-1$. 这个运动类似于随机游动, 只不过其转移是出现在随机时刻, 而不是固定的周期时间.

构造生灭过程的传统方法是预先指定生与灭的参数 $\{\lambda_i, \mu_i\}_{i=0}^{\infty}$, 然后使用上述关于等候时间和到达各状态的条件转移概率建立轨道结构. 我们确定过程的实现可如下进行. 设 $X(0) = i$; 粒子在状态 i 逗留一个随机长度时间 (以 $\lambda_i + \mu_i$ 为参数的指数分布), 然后以概率 $\lambda_i/(\mu_i + \lambda_i)$ 运动到 $i+1$, 以概率 $\mu_i/(\lambda_i + \mu_i)$ 运动到 $i-1$. 接着, 粒子又在新的状态逗留一个随机长度时间后转移到与它相邻的状

态之一, 如此反复. 更具体地, 我们从参数为 $(\mu_i+\lambda_i)$ 的指数分布中观测到一个值 t_1, 这样确定初始逗留状态 i 的时间. 然后我们投掷一个硬币, 它出现正面的概率为 $p_i=\lambda_i/(\lambda_i+\mu_i)$, 如果正面 (反面) 出现, 我们把粒子移动到状态 $i+1(i-1)$. 在状态 $i+1$, 我们从参数为 $(\mu_{i+1}+\lambda_{i+1})$ 的指数分布中观测到一个值 t_2, 以此确定粒子在第二个状态逗留的时间. 如果粒子在第一次转移进入状态 $i-1$, 则随后逗留时间 t_2' 是从参数为 $(\lambda_{i-1}+\mu_{i-1})$ 的指数分布中进行一次观测得到的. 在第二次等候之后, 如同前面一样, 进行一次二项试验以选择下一次所进入的状态, 等等.

上述抽样过程所得到一个典型结果决定了过程的一个实现. 它的形式是

$$X(t)=\begin{cases} i & 0<t<t_1, \\ i+1 & t_1<t<t_1+t_2, \\ i & t_1+t_2<t<t_1+t_2+t_3, \\ \vdots & \\ . & \end{cases}$$

这样, 通过从适当的指数分布和二项分布中抽样, 构造了过程的一个典型样本轨道. 现在我们有可能以某种方式规定这些轨道 (过程的实现) 集合的概率测度, 使 $P_{ij}(t)$ 满足 (4.2) 和 (4.3) 以及无穷小关系式. 这个结果相当深入, 它的严格讨论已超过本书讨论范围, 以这种方式得到的过程称为与矩阵 $\boldsymbol{A}$ 相联系的极小过程. 134

极小过程的上述构造法是十分重要的, 因为无穷小参数确定的满足 (4.2)、(4.3) 及本书 A. 中的假定的随机过程不必唯一. 事实上, 可能有几个马尔可夫过程具有相同的无穷小生成元. 这个问题比较复杂, 我们推荐读者参看 Chung 的书 [1]. 对于生灭过程这一特殊情况, 具有满足 (4.2)、(4.3) 和无穷小关系式的转移概率函数 $P_{ij}(t)$ 的马氏过程存在唯一的充分条件为

$$\sum_{n=0}^{\infty}\pi_n\sum_{k=0}^{n}\frac{1}{\lambda_k\pi_k}=\infty, \tag{4.5}$$

其中 $\pi_0=1,\pi_n=\dfrac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n},\ n=1,2,\cdots$.

在大多数生灭过程实际例子中, 式 (4.5) 能满足, 因此与相应参数联系的生灭过程是唯一确定的.

4.5 生灭过程的微分方程

与纯生过程及泊松过程情况类似, 转移概率 $P_{ij}(t)$ 满足著名的**向后科尔莫戈罗夫微分方程**, 即

1. K. L. Chung, *Markov chains with stationary transition probabilities*, Springer-Verlag, Berlin, 1960.

$$P'_{0j}(t) = -\lambda_0 P_{0j}(t) + \lambda_0 P_{1j}(t),$$
$$P'_{ij}(t) = \mu_i P_{i-1,j}(t) - (\lambda_i + \mu_i)P_{ij}(t) + \lambda_i P_{i+1,j}(t), \quad i \geqslant 1, \tag{5.1}$$

且边界条件为 $P_{ij}(0) = \delta_{ij}$.

为了导出这些式子, 由方程 (4.3) 我们有

135

$$\begin{aligned} P_{ij}(t+h) &= \sum_{k=0}^{\infty} P_{ik}(h)P_{kj}(t) \\ &= P_{i,i-1}(h)P_{i-1,j}(t) + P_{i,i}(h)P_{ij}(t) + P_{i,i+1}(h)P_{i+1,j}(t) \\ &\quad + \sum_k{}' P_{ik}(h)P_{kj}(t), \end{aligned} \tag{5.2}$$

其中最后一个和号是对所有 $k \neq i-1, i, i+1$ 求和的. 利用 4.4 节假设 1, 2 和 3, 我们得

$$\begin{aligned} \sum_k{}' P_{ik}(h)P_{kj}(t) &\leqslant \sum_k{}' P_{ik}(h) \\ &= 1 - [P_{i,i}(h) + P_{i,i-1}(h) + P_{i,i+1}(h)] \\ &= 1 - [1 - (\lambda_i + \mu_i)h + o(h) + \mu_i h + o(h) + \lambda_i h + o(h)] \\ &= o(h), \end{aligned}$$

所以

$$P_{ij}(t+h) = \mu_i h P_{i-1,j}(t) + (1 - (\lambda_i + \mu_i)h)P_{ij}(t) + \lambda_i h P_{i+1,j}(t) + o(h).$$

把 $P_{ij}(t)$ 移到左边并用 h 除之, 然后令 $h \downarrow 0$, 得

$$P'_{ij}(t) = \mu_i P_{i-1,j}(t) - (\lambda_i + \mu_i)P_{ij}(t) + \lambda_i P_{i+1,j}(t).$$

以上分析只是第 14 章向后微分方程式推导的特例.

向后方程是由分解时间区间 $(0, t+h)$ 为两段区间: $(0, h)$ 和 $(h, t+h)$, 其中 h 为较小的正数, 然后分别检验在每段区间上的转移而推导出来的.

方程 (5.1) 的特点是初始状态是可以变化的.

把时间区间 $(0, t+h)$ 分为 $(0, t)$ 和 $(t, t+h)$ 两段, 并采用上面分析方法, 在更加严格条件之下, 我们可导出另一个微分方程组

$$P'_{i0}(t) = -\lambda_0 P_{i0}(t) + \mu_1 P_{i1}(t),$$
$$P'_{ij}(t) = \lambda_{j-1} P_{i,j-1}(t) - (\lambda_j + \mu_j)P_{ij}(t) + \mu_{j+1} P_{i,j+1}(t), j \geqslant 1, \tag{5.3}$$

伴有相同的初始条件 $P_{ij}(0) = \delta_{ij}$. 这就是著名的**向前科尔莫戈罗夫微分方程式**. 为此, 我们在方程 (5.2) 中调换 t 和 h, 除满足假设 1,2 和 3 之外在更强的条件下, 可

以证明最后一项也是 $o(h)$. 其余的论证和前面一样. 这些微分方程的作用将在下面讨论的例子中显示出来.

(5.3) 成立的一个充分条件是对于 $k \neq j,\ j-1,\ j+1$ 有 $(P_{kj}(h))/h = o(1)$, 其中 $o(1)$ 表示当 $h \to 0$ 时趋于 0, 并对固定的 j 关于 k 是一致有界的. 在这种情况下, 容易证明 $\sum_k {}' P_{ik}(t)P_{kj}(h) = o(h)$.

136

在介绍下面例子之前, 我们首先简要讨论 $P_{ij}(t)$ 当 t 变得足够大时的性质. 可以证明下面的极限

$$\lim_{t\to\infty} P_{ij}(t) = p_j \tag{5.4}$$

存在且与初始状态 i 无关, 并且它们满足

$$\begin{aligned} &-\lambda_0 p_0 + \mu_1 p_1 = 0, \\ &\lambda_{j-1}p_{j-1} - (\lambda_j + \mu_j)p_j + u_{j+1}p_{j+1} = 0, \quad j \geqslant 1. \end{aligned} \tag{5.5}$$

这些方程相当于把 (5.3) 的左边部分置为 0. 由于 $\sum_j P_{ij}(t) = 1$, 可断定 $\sum_j p_j$ 收敛. 如果 $\sum_j p_j = 1$, 则序列 $\{p_j\}$ 称为 "平稳分布", 这是因为 p_j 也满足

$$p_j = \sum_{i=0}^{\infty} p_i P_{ij}(t), \tag{5.6}$$

此式告诉我们, 如果过程以概率 p_i 从状态 i 出发, 则在任意给定时刻 t, 它将以相同的概率 p_i 处于状态 i. (5.6) 的证明可在 (4.3) 和 (5.4) 中令 $t \uparrow \infty$ 且利用 $\sum_{i=0}^{\infty} p_i < \infty$ 得到. 由归纳法可得到 (5.5) 的解. 令

$$\pi_0 = 1,\ \pi_j = \frac{\lambda_0\lambda_1\cdots\lambda_{j-1}}{\mu_1 u_2 \cdots \mu_j}, \quad j \geqslant 1,$$

则有 $p_1 = \lambda_0\mu_1^{-1}p_0 = \pi_1 p_0$. 假设对于 $k = 1, \cdots, j$ 有 $p_k = \pi_k p_0$, 我们有

$$\begin{aligned} \mu_{j+1}p_{j+1} &= (\lambda_j + \mu_j)\pi_j p_0 - \lambda_{j-1}\pi_{j-1}p_0 \\ &= \lambda_j\pi_j p_0 + (\mu_j\pi_j - \lambda_{j-1}\pi_{j-1})p_0 \\ &= \lambda_j\pi_j p_0, \end{aligned}$$

最后得 $p_{j+1} = \pi_{j+1}p_0$. 为了使序列 $\{p_j\}$ 是一个分布, 必须成立 $\sum p_j = 1$. 若 $\sum \pi_k < \infty$, 我们有

$$p_i = \frac{\pi_j}{\sum \pi_k} \quad j = 0, 1, 2, \cdots.$$

若 $\sum \pi_k = \infty$, 则必定有 $p_0 = 0$, 因而所有 p_j 都等于 0, 此时不存在极限平稳分布.

4.6 生灭过程的例子

例 1 有迁入的线性增长 生灭过程称为线性增长过程, 如果 $\lambda_n = \lambda n + a$ 和 [137] $\mu_n = \mu n$, 其中 $\lambda > 0$, $\mu > 0$, $a > 0$. 这样的过程出现在研究生物繁殖和群体增长中. 如果状态 n 描述现在生物群体总量, 则平均瞬时增长率为 $\lambda n + a$. 类似地, 经过一小段时间之后过程的状态减少 1 个的概率是 $\mu nt + o(t)$. 因子 λn 表示群体总量由当时状态所引起的自然增长, 而第二个分量 a 可解释为由于外来因素 (如迁入) 而引起总量增长. 显然, 分量 μn 表示当前生物群体的平均无穷小死亡率.

如果我们把上述 λ_n 和 μ_n 的值代入 (5.3), 则得到

$$
\begin{aligned}
&P'_{i0}(t) = -aP_{i0}(t) + \mu P_{i1}(t),\\
&P'_{ij}(t) = (\lambda(j-1)+a)P_{i,j-1}(t) - ((\lambda+\mu)j+a)P_{ij}(t) + \mu(j+1)P_{i,j+1}(t), j \geqslant 1.
\end{aligned}
$$

我们用 j 乘以第 j 个方程并且求和, 推得

$$E[X(t)] = M(t) = \sum_{j=1}^{\infty} jP_{ij}(t)$$

满足微分方程

$$
\begin{cases}
M'(t) = a + (\lambda - \mu)M(t), \\
M(0) = i, \qquad\qquad \text{如果 } X(0) = i.
\end{cases}
$$

这个方程的解是

$$M(t) = at + i \quad \text{若 } \lambda = \mu$$

且

$$M(t) = \frac{a}{\lambda - \mu}\{e^{(\lambda-\mu)t} - 1\} + ie^{(\lambda-\mu)t} \quad \text{若} \lambda \neq \mu.$$

其二阶矩或方差可用类似方法求得. 一个有趣的事实是若 $\lambda \geqslant \mu$, 则当 $t \to \infty$ 时, $M(t) \to \infty$; 而若 $\lambda < \mu$, 则对足够大的 t 平均群体总量近似于

$$\frac{a}{\mu - \lambda}.$$

例 2 排队 排队过程是描述顾客们到达某指定位置并得到某种服务的一个过程. 例如在银行出纳员窗口或者在超级市场的缴款处都会出现这一类现象. 我们假定顾客先后到达的间隔时间和对于某个顾客提供服务的时间服从某些概率分布. [138] 在给定时刻 t 排队长度用 $X(t)$ 表示.

在生灭过程中若对所有 i 令 $\lambda_i=\lambda$, 则此过程即是连续时间排队过程的简单情况. 这时, 系统的状态是排队长度, 前后两个顾客到达的间隔时间是服从参数为 λ 的指数分布的独立随机变量, 为单个顾客服务的时间服从参数为 μ_n(依赖于当时排队长度) 的指数分布. 在每次服务完成之后, 排队长度减少 1, 并且每个新顾客到达又使排队长度增加 1. 具有单一服务的排队过程古典情况对应于 $\mu_i=\mu,\ i\geqslant 1$, 即每次服务时间都服从相同参数 μ 的指数分布, 而与排队长度无关.

古典的电话系统模型可以作为具有无限多服务的排队生灭过程, 每次服务时间的分布有相同的参数 μ, 因此 $\mu_i=i\mu,\ i\geqslant 1$. 此式的依据如下: 假设队伍由 i 个顾客组成, 由于有无限多服务员, 因此所有顾客可以同时接受服务. 又因为每个顾客的服务时间长度是相互独立的且服从参数为 μ 的指数分布. 由此推得至少有一个顾客完成服务 (即排队长度至少减 1) 的时间也是服从指数分布的, 但其参数为 $i\mu$(读者可证明此事实).

除了上述两个特殊情况之外, 我们还可以通过适当给定参数 μ_k 来讨论许多其他的排队模型. 例如, 具有 n 个服务, 每个服务时间具有相同参数 μ 的指数分布的排队模型. 此时, 当 $1\leqslant k\leqslant n$ 时 $\mu_k=k\mu$, 而当 $i\geqslant n$ 时 $\mu_i=n\mu$.

对于单一服务的排队过程, 若 $\lambda<\mu$, 其平稳分布容易求得. 事实上, 在这种情况下有

$$\pi_n=\frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}=\left(\frac{\lambda}{\mu}\right)^n,$$

由此推得

$$p_n=\frac{\mu-\lambda}{\mu}\left(\frac{\lambda}{\mu}\right)^n,\quad n\geqslant 0,$$

即具有均值为 $\lambda(\mu-\lambda)^{-1}$ 的几何分布. 139

上式使我们能解释带有平稳性的许多问题. 若过程已经过很长时间, 且 $\lambda<\mu$, 则马上被服务的概率是

$$p_0=\left(1-\frac{\lambda}{\mu}\right).$$

我们还可以求出当 $\lambda<\mu$ 情况下等候时间的平稳分布. 若一个新来顾客发现有 n 个人排在他的前面, 他的全部等候时间 T 是他自己被服务的时间和他前面的人被服务的时间之和, 每个服务时间都服从参数为 μ 的指数分布, 并且服务时间与排队长度无关. 这样 T 具有阶数为 $n+1$, 参数为 μ 的 Γ 分布:

$$\Pr\{T\leqslant t|\text{有 } n \text{ 个人在前面}\}=\int_0^t\frac{\mu^{n+1}\tau^n\mathrm{e}^{-\mu\tau}}{\Gamma(n+1)}\mathrm{d}\tau. \tag{6.1}$$

由全概率公式, 我们有

$$\Pr\{T\leqslant t\}=\sum_{n=0}^{\infty}\Pr\{T\leqslant t|\text{有 } n \text{ 个人在前面}\}\cdot\left(\frac{\lambda}{\mu}\right)^n\left(1-\frac{\lambda}{\mu}\right),$$

因为 $(\lambda/\mu)^n(1-\lambda/\mu)$ 是在平稳情况下一个顾客到达并发现队伍前头有 n 个人的概率. 将 (6.1) 代入, 得到

$$\begin{aligned}\Pr\{T \leqslant t\} &= \sum_{n=0}^{\infty}\int_0^t \frac{\mu^{n+1}\tau^n \mathrm{e}^{-\mu\tau}}{\Gamma(n+1)}\left(\frac{\lambda}{\mu}\right)^n\left(1-\frac{\lambda}{\mu}\right)\mathrm{d}\tau \\ &= \int_0^t \mu\mathrm{e}^{-\mu\tau}\left(1-\frac{\lambda}{\mu}\right)\sum_{n=0}^{\infty}\frac{\tau^n\lambda^n}{\Gamma(n+1)}\mathrm{d}\tau \\ &= \int_0^t\left(1-\frac{\lambda}{\mu}\right)\mu\exp\left\{-\tau\mu\left(1-\frac{\lambda}{\mu}\right)\right\}\mathrm{d}\tau \\ &= 1-\exp\left[-t\mu\left(1-\frac{\lambda}{\mu}\right)\right],\end{aligned}$$

这也是一个指数分布.

关于非平稳的问题, 其本质是对所有 t 确定 $P_{ij}(t)$. 这是一个比较难的问题, 但它已被解决. 阐述这个问题解决的细节已超出本书目的, 我们建议有兴趣读者参阅列在本章末关于排队论的任何一本高级著作.

140

对电话系统问题, 当 $\lambda_n = \lambda$ 和 $\mu_n = n\mu$ 时, 易见

$$p_n = \frac{\mathrm{e}^{-\lambda/\mu}(\lambda/\mu)^n}{n!},$$

这是熟知的均值为 λ/μ 的泊松分布, 如例 1 一样, 容易验证

$$M(t) = \sum_{j=0}^{\infty} jP_{ij}(t)$$

满足方程

$$M'(t) = \lambda - \mu M(t),$$

它的解是

$$M(t) = \frac{\lambda}{\mu}\left(1-\mathrm{e}^{-\mu t}\right) + i\mathrm{e}^{-\mu t}.$$

若我们令 $t \to \infty$, 则 $M(t) \to \lambda/\mu$, 这是上述给出的平稳分布的均值.

例 3　某些遗传模型　考虑由 N 个个体组成群体, 这些个体或者具有 a 型基因或者具有 A 型基因. 过程 $X(t)$ 的状态表示在时刻 t 的 a 型个体数目. 我们假设在时间区间 $(t, t+h)$ 状态改变的概率是 $\lambda h + o(h)$, 它与 $X(t)$ 的值无关, 并且在时间区间 h 内出现两个或更多个改变的概率是 $o(h)$.

群体结构的改变由如下方式实现, 一个个体将被从群体中随机选择的另一个所代替; 即, 若 $X(t) = j$, 则 a 型被选择并用以替换的概率为 j/N, 而 A 型的为 $1-j/N$.

我们称这为灭过程，而新生个体，按照下列方式产生，从群体中随机地选取一个来决定新个体的类型，以此新个体来代替死亡的个体. 该模型中引入了变异的因素，变异使得新个体的类型在出生时可能会相互变化. 具体地，令 γ_1 表示 a 型变异为 A 型的概率，γ_2 表示 A 型变异为 a 型的概率.

这样，加入群体的新个体属于 a 型的概率是

$$\frac{j}{N}(1-\gamma_1)+\left(1-\frac{j}{N}\right)\gamma_2 \tag{6.2}$$

我们推导此公式如下：我们选择一个 a 型且没有发生变异的概率是 $(j/N)(1-\gamma_1)$. 141 而且，如果我们挑选到一个 A 型的，随后可能变异为 a 型的，这个事件的概率为 $(1-j/N)\gamma_2$. 这两个概率相加即得 (6.2)

我们断言在发生状态改变时，$X(t+)-X(t)=1$ 的条件概率是

$$\left(1-\frac{j}{N}\right)\left[\frac{j}{N}(1-\gamma_1)+\left(1-\frac{j}{N}\right)\gamma_2\right],\quad 其中\ X(t)=j. \tag{6.3}$$

事实上，群体中 a 型个体数量只有在一个 A 型个体死亡后 (从而被代替) 才能增加. 这个概率是 $1-(j/N)$. 第二个因子表示新个体为 a 型的概率，如 (6.2) 所示.

用类似方法求得在发生状态改变时 $X(t+)-X(t)=-1$ 的条件概率是

$$\frac{j}{N}\left[\left(1-\frac{j}{N}\right)(1-\gamma_2)+\frac{j}{N}\gamma_1\right],\quad 其中\ X(t)=j.$$

这样，上述的随机过程是一个具有有限状态的生灭过程 [1]，对应于 a 型数量为 $j(0\leqslant j\leqslant N)$ 的无穷小生和灭的比率分别是

$$\lambda_j=\lambda\left(1-\frac{j}{N}\right)\left[\frac{j}{N}(1-\gamma_1)+\left(1-\frac{j}{N}\right)\gamma_2\right]$$

和

$$u_j=\lambda\frac{j}{N}\left[\frac{j}{N}\gamma_1+\left(1-\frac{j}{N}\right)(1-\gamma_2)\right].$$

虽然这些参数似乎很复杂，但若我们令群体总量 $N\to\infty$，且每个个体变异的概率 γ_1,γ_2 依如下方式趋于 0：$\gamma_1 N\to k_1$ 和 $\gamma_2 N\to k_2$，其中 $0<k_1,k_2<\infty$，此时平稳测度 $\{\pi_k\}_{k=0}^N$ 究竟如何，却是一个令人感兴趣的问题. 为方便，我们把过程的状态转换到区间 [0,1]，这只需定义新的状态为 j/N，即 a 型个体在群体中的比例. 142 为了研究在一个固定分数 $x(0<x<1)$ 上的平稳密度，我们将求 π_k 与 $k=[xN]$ 且 $k\to\infty$ 时极限，其中 $[xN]$ 表示不超过 xN 的最大整数.

1. 生灭过程的定义是对无限状态给出的，而对于仅具有有限状态的生灭过程的定义和分析是比较简单的，留给读者进行.

改写 λ_j 和 μ_j 如下

$$\lambda_j = \frac{\lambda(N-j)}{N^2}(1-\gamma_1-\gamma_2)j\left(1+\frac{a}{j}\right), \quad \text{其中 } a = \frac{N\gamma_2}{1-\gamma_1-\gamma_2}$$

和

$$\mu_j = \frac{\lambda(N-j)}{N^2}(1-\gamma_1-\gamma_2)j\left(1+\frac{b}{N-j}\right), \quad \text{其中 } b = \frac{N\gamma_1}{1-\gamma_1-\gamma_2}$$

则

$$\begin{aligned}\ln \pi_k &= \sum_{j=0}^{k-1} \ln \lambda_j - \sum_{j=1}^{k} \ln \mu_j \\ &= \sum_{j=1}^{k-1} \ln\left(1+\frac{a}{j}\right) - \sum_{j=1}^{k-1} \ln\left(1+\frac{b}{N-j}\right) + \ln Na \\ &\quad - \ln(N-k)k\left(1+\frac{b}{N-k}\right).\end{aligned}$$

利用表达式

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots, \quad |x| < 1,$$

可以得到

$$\sum_{j=1}^{k-1} \ln\left(1+\frac{a}{j}\right) = a\sum_{j=1}^{k-1}\frac{1}{j} + c_k,$$

其中当 $k \to \infty$ 时 c_k 趋于有限极限. 因此利用关系式

$$\sum_{j=1}^{k-1}\frac{1}{j} \sim \ln k, \quad \text{当 } k \to \infty \text{ 时},$$

我们有

$$\sum_{j=1}^{k-1} \ln\left(1+\frac{a}{j}\right) \sim \ln k^a + c_k, \quad \text{当 } k \to \infty \text{ 时}.$$

以类似方法我们得到

143

$$\sum_{j=1}^{k-1} \ln\left(1+\frac{b}{N-j}\right) \sim \ln\frac{N^b}{(N-k)^b} + d_k, \quad \text{当 } k \to \infty \text{ 时},$$

其中当 $k \to \infty$ 时 d_k 趋于一个有限极限. 利用上述关系式, 我们有

$$\ln \pi_k \sim \ln\left(C_k \frac{k^a (N-k)^b Na}{N^b (N-k)k}\right), \quad \text{当 } k \to \infty \text{ 时}, \tag{6.4}$$

其中 $\ln C_k = c_k + d_k$ 当 $k \to \infty$ 时趋于一个极限, 假设是 C. 注意到当 $N \to \infty$ 时 $a \to \kappa_2$ 和 $b \to \kappa_1$, 并由于 $k = [Nx]$, 当 $N \to \infty$ 时我们有

$$\pi_k \sim C\kappa_2 N^{\kappa_2 - 1} x^{\kappa_2 - 1}(1-x)^{\kappa_1 - 1}.$$

由 (6.4) 得

$$\pi_k \sim aC_k k^{a-1}\Big(1 - \frac{k}{N}\Big)^{b-1}.$$

所以

$$\frac{1}{N^a}\sum_{k=0}^{N-1}\pi_k \sim \frac{a}{N}\sum_{k=0}^{N-1} C_k\Big(\frac{k}{N}\Big)^{a-1}\Big(1-\frac{k}{N}\Big)^{b-1}.$$

由于当 $k \to \infty$ 时 $C_k \to C$, 上式右边部分可作为如下积分

$$\kappa_2 C\int_0^1 x^{\kappa_2 - 1}(1-x)^{k_1 - 1}\mathrm{d}x$$

的黎曼和, 这样

$$\sum_{i=0}^{N}\pi_i \sim N^{\kappa_2}\kappa_2 C\int_0^1 x^{\kappa_2-1}(1-x)^{\kappa_1-1}\mathrm{d}x,$$

所以在 $[0,1]$ 是密度函数为

$$\frac{\pi_k}{\sum \pi_i} \sim \frac{1}{N}\frac{x^{\kappa_2-1}(1-x)^{\kappa_1-1}}{\displaystyle\int_0^1 x^{\kappa_2-1}(1-x)^{\kappa_1-1}\mathrm{d}x} = \frac{x^{\kappa_2-1}(1-x)^{\kappa_1-1}\mathrm{d}x}{\displaystyle\int_0^1 x^{\kappa_2-1}(1-x)^{\kappa_1-1}\mathrm{d}x},$$

此处 $\mathrm{d}x \sim 1/N$. 这是一个参数为 κ_1 和 κ_2 的 Beta 分布.

例 4 Logistic 过程 我们考虑一个群体, 其总量 $X(t)$ 的值域对于所有 $t \geqslant 0$ 是在两个固定整数 N_1 和 $N_2 (N_1 < N_2)$ 之间. 假设每个个体在时刻 t 生和灭比率分别为

$$\lambda = \alpha(N_2 - X(t)) \quad 和 \quad \mu = \beta(X(t) - N_1),$$

并且群体中每个成员的行动彼此独立. 因此该群体的生和灭比率分别为

$$\lambda_n = \alpha n(N_2 - n) \quad 和 \quad \mu_n = \beta n(n - N_1).$$

[144]

这是因为若群体总量 $X(t)$ 是 n, 则 n 个个体的每一个都有一个无穷小生的比率 λ, 所以 $\lambda_n = \alpha n(N_2 - n)$. 同样的理由可用于解释 μ_n.

在这样条件之下, 可以期望生过程的值是在两个常数 N_1 和 N_2 之间波动. 因为如果 $X(t)$ 靠近 N_2, 此时灭的比率高, 生的比率低, 则 $X(t)$ 将朝 N_1 移动. 最后, 过程将表现为在两个极限 N_1 和 N_2 之间平稳的波动.

在这种情况下. 平稳分布是

$$p_{N_1+m}=\frac{c}{N_1+m}\binom{N_2-N_1}{m}\left(\frac{\alpha}{\beta}\right)^m,\quad m=0,1,2,\cdots,N_2-N_1,$$

其中 c 是一个适当常数, 以保证 $\sum_m p_{N_1+m}=1$. 为此, 我们只要注意

$$\begin{aligned}\pi_{N_1+m}&=\frac{\lambda_{N_1}\lambda_{N_1+1}\cdots\lambda_{N_1+m}}{\mu_{N_1+1}\mu_{N_1+2}\cdots\mu_{N_1+m}}\\&=\frac{\alpha^m N_1(N_1+1)\cdots(N_1+m-1)(N_2-N_1)\cdots(N_2-N_1-m+1)}{\beta^m(N_1+1)\cdots(N_1+m)m!}\\&=\frac{N_1}{N_1+m}\binom{N_2-N_1}{m}\left(\frac{\alpha}{\beta}\right)^m.\end{aligned}$$

4.7 带有吸收状态的生灭过程

生灭过程中当 $\lambda_0=0$ 时的情形是非常重要的. 这时状态 0 成为吸收状态. 当粒子从状态 1 开始转移时, 将以概率 $\lambda_1/(\lambda_1+\mu_1)$ 运动到状态 2, 以概率 $\mu_1/(\lambda_1+\mu_1)$ 陷入状态 0. 以 0 为吸收状态的生灭过程的一个重要例子是没有迁入的线性增长过程 (4.6 节中的例 1). 在这种情况下, $\lambda_n=n\lambda$ 和 $\mu_n=n\mu$. 由于群体的增长只由当时的群体总量决定, 因此当群体总量为 0 时, 此后它就一直停留在 0, 即 0 是吸收状态.

A. 被状态 0 吸收的概率

计算从状态 $i(i\geqslant 1)$ 出发最后被状态 0 吸收的概率将是一个有趣的问题. 按推测这个事件不是必然事件, 因为可以想象粒子 (即状态变量) 可能永远游动在状态 (1,2,$\cdots$) 或可能趋于无穷.

145

设 $\mu_i(i=1,2,\cdots)$ 表示从初始状态 i 出发被状态 0 吸收的概率. 通过考虑第一次转移之后所有可能状态, 我们可以写出 μ_i 的递推公式. 由于第一次转移必然是

$$i\longrightarrow i+1\quad \text{以概率}\frac{\lambda_i}{\mu_i+\lambda_i}$$

$$i\longrightarrow i-1\quad \text{以概率}\frac{\mu_i}{\mu_i+\lambda_i}$$

我们直接得到

$$u_i=\frac{\lambda_i}{\mu_i+\lambda_i}u_{i+1}+\frac{\mu_i}{\mu_i+\lambda_i}u_{i-1}\quad i\geqslant 1,\tag{7.1}$$

其中 $u_0=1$. 导出 (7.1) 的另一个方法是考虑与生灭过程相联系的 "嵌入随机游动". 具体地说, 我们仅仅在转移时刻考察生灭过程. 以这种方式产生的离散马尔可夫链

用 $\{Y_n\}_{n=0}^{\infty}$ 表示, 其中 $Y_0 = X_0$ 是初始状态, $Y_n(n \geqslant 1)$ 是第 n 次转移的状态. 显然, 转移概率矩阵具有形式

$$\boldsymbol{P} = \left\| \begin{matrix} 1 & 0 & 0 & 0 & \cdots \\ q_1 & 0 & p_1 & 0 & \cdots \\ 0 & q_2 & 0 & p_2 & \cdots \\ \vdots & \vdots & & & \end{matrix} \right\|,$$

其中

$$p_i = \frac{\lambda_i}{\lambda_i + \mu_i} = 1 - q_i \quad (i \geqslant 1).$$

嵌入随机游动被状态 0 吸收的概率和对应的生灭过程是相同的, 因为这两个过程有相同的转移.

现在我们在 $u_0 = 1$ 和 $0 \leqslant u_i \leqslant 1 (i \geqslant 1)$ 的条件下来解 (7.1). 改写 (7.1) 为

$$(u_{i+1} - u_i) = \frac{\mu_i}{\lambda_i}(u_i - u_{i-1}), \quad i \geqslant 1.$$

置 $v_i = u_{i+1} - u_i$, 我们得到

$$v_i = \frac{\mu_i}{\lambda_i} v_{i-1}, \quad i \geqslant 1.$$

146

由最后关系式的迭代得到公式

$$u_{i+1} - u_i = v_i = \left(\prod_{j=1}^{i} \frac{\mu_j}{\lambda_j} \right) v_0, \quad i \geqslant 1.$$

对这些方程从 $i = 1$ 到 $i = m$ 求和, 我们有

$$u_{m+1} - u_1 = (u_1 - 1) \sum_{i=1}^{m} \left(\prod_{j=1}^{i} \frac{\mu_j}{\lambda_j} \right), \quad m \geqslant 1 \tag{7.2}$$

因为 u_m 是概率, 且以 1 为上界, 可知如果

$$\sum_{i=1}^{\infty} \left(\prod_{j=1}^{i} \frac{\mu_j}{\lambda_j} \right) = \infty \tag{7.3}$$

则一定有 $u_1 = 1$ 和 $u_m = 1, m \geqslant 2$. 换句话说, 若 (7.3) 成立, 则从任何初始状态出发最后被状态 0 吸收是肯定的. 假设 $0 < u_1 < 1$, 则当然有

$$\sum_{i=1}^{\infty} \left(\prod_{j=1}^{i} \frac{\mu_j}{\lambda_j} \right) < \infty.$$

显然, u_m 是随着 m 增大而减小的, 因为从状态 m 到状态 0 的过渡期间需要经过中间状态. 更进一步, 我们证明 $u_m \to 0(m \to \infty)$. 假设相反, 即 $u_m \geqslant \alpha > 0(m \geqslant 1)$, 由简单概率论命题推得 $u_m \equiv 1(m \geqslant 1)$. (读者可补充出其正式证明.) 在 (7.2) 中令 $m \to \infty$, 解得

$$u_1 = \frac{\sum_{i=1}^{\infty}\left(\prod_{j=1}^{i} \mu_j/\lambda_j\right)}{1+\sum_{i=1}^{\infty}\left(\prod_{j=1}^{i} \mu_j/\lambda_j\right)}.$$

此外, 我们有

$$u_{m+1} = \frac{\sum_{i=m+1}^{\infty}\left(\prod_{j=1}^{i} \mu_j/\lambda_j\right)}{1+\sum_{i=1}^{\infty}\left(\prod_{j=1}^{i} \mu_j/\lambda_j\right)}, \quad m \geqslant 1.$$

在线性增长生灭过程的例子中, 由于 $\mu_n = n\mu$, $\lambda_n = n\lambda$, 由直接计算可得

147

$$\begin{aligned} u_m &= \left(\frac{\mu}{\lambda}\right)^m \quad \text{当 } \mu < \lambda \\ u_m &= 1 \qquad\quad\ \ \text{当 } \mu \geqslant \lambda \end{aligned} \quad (m \geqslant 1).$$

B. 平均吸收时间

考虑这样一个问题：如何确定从状态 m 出发到达吸收状态的平均时间.

我们假设条件 (7.3) 成立, 因此吸收是肯定的. 注意, 我们不能把此问题归结为考虑嵌入随机游动, 因为在每个状态的实际逗留时间与计算平均吸收时间是有关的.

设 ω_i 是从状态 i 出发的平均吸收时间 (它可能是无限的). 考虑在第一次转移之后的可能状态以及利用在状态 i 的平均等候时间为 $(\lambda_i+\mu_i)^{-1}$ 这个事实 (实际上是以 $\lambda_i+\mu_i$ 为参数的指数分布), 我们导出递推关系式

$$\omega_i = \frac{1}{\lambda_i+\mu_i} + \frac{\lambda_i}{\lambda_i+\mu_i}\omega_{i+1} + \frac{\mu_i}{\lambda_i+\mu_i}\omega_{i-1}, \quad i \geqslant 1, \tag{7.4}$$

其中约定 $\omega_0 = 0$. 令 $z_i = \omega_i - \omega_{i+1}$, 并改写 (7.4) 为

$$z_i = \frac{1}{\lambda_i} + \frac{\mu_i}{\lambda_i} z_{i-1}, \quad i \geqslant 1. \tag{7.5}$$

由迭代得

$$z_m = \frac{1}{\lambda_m} + \frac{\mu_m}{\lambda_m}\frac{1}{\lambda_{m-1}} + \frac{\mu_m\mu_{m-1}}{\lambda_m\lambda_{m-1}} z_{m-2},$$

最后,

$$z_m = \sum_{i=1}^{m} \frac{1}{\lambda_i} \prod_{j=i+1}^{m} \frac{\mu_j}{\lambda_j} + \left(\prod_{j=1}^{m} \frac{\mu_j}{\lambda_j}\right) z_0.$$

$\left(\text{乘积} \prod_{m+1}^{m} \mu_j/\lambda_j \text{ 定义为 } 1.\right)$

把 z_m 换回 ω_m, 我们有

$$\omega_m - \omega_{m+1} = \sum_{i=1}^{m} \frac{1}{\lambda_i} \prod_{j=i+1}^{m} \frac{\mu_j}{\lambda_j} - \omega_1 \prod_{j=1}^{m} \frac{\mu_j}{\lambda_j}, \quad m \geqslant 1. \tag{7.6}$$

引用记号

$$\rho_i = \frac{\lambda_1 \lambda_2 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i},$$

则有

$$\sum_{i=1}^{m} \frac{1}{\lambda_i} \prod_{j=i+1}^{m} \frac{\mu_j}{\lambda_j} = \prod_{j=1}^{m} \frac{\mu_j}{\lambda_j} \sum_{i=1}^{m} \rho_i. \tag{7.7}$$

148

然后利用 (7.7), 关系式 (7.6) 成为

$$\left(\prod_{j=1}^{m} \frac{\lambda_j}{\mu_j}\right)(\omega_m - \omega_{m+1}) = \sum_{i=1}^{m} \rho_i - \omega_1. \tag{7.8}$$

注意, 如果 $\sum_{i=1}^{\infty} \rho_i = \infty$, 由 (7.8) 知必定有 $\omega_1 = \infty$. 事实上, 由于 $\omega_m < \omega_{m+1}$ 对于所有的 m 成立, 若我们假设 ω_1 是有限的, 则当 m 足够大时, 这个性质将不成立.

现在我们假设 $\sum_{i=1}^{\infty} \rho_i < \infty$; 在 (7.8) 中令 $m \to \infty$ 得

$$\omega_1 = \sum_{i=1}^{\infty} \rho_i - \lim_{m \to \infty} \left(\prod_{j=1}^{m} \frac{\lambda_j}{\mu_j}\right)(\omega_m - \omega_{m+1}).$$

虽然比较麻烦但仍可以证明

$$\lim_{m \to \infty} \left(\prod_{j=1}^{m} \frac{\lambda_j}{\mu_j}\right)(\omega_m - \omega_{m+1}) = 0,$$

因此,

$$\omega_1 = \sum_{i=1}^{\infty} \rho_i.$$

我们把这节的讨论总结为如下的定理:

定理 7.1　考虑一个生和灭参数分别为 $\lambda_n, \mu_n, n \geqslant 1$ 的生灭过程, 其中 $\lambda_0 = 0$, 因此 0 是吸收状态. 从初始状态 m 出发被状态 0 吸收的概率是

$$\begin{cases} \dfrac{\sum\limits_{i=m}^{\infty}\Big(\sum\limits_{j=1}^{i}\mu_j/\lambda_j\Big)}{1+\sum\limits_{i=1}^{\infty}\Big(\prod\limits_{j=1}^{i}\mu_j/\lambda_j\Big)} & \text{若 } \sum\limits_{i=1}^{\infty}\Big(\prod\limits_{j=1}^{i}\dfrac{\mu_j}{\lambda_j}\Big) < \infty, \\ 1 & \text{若 } \sum\limits_{i=1}^{\infty}\Big(\prod\limits_{j=1}^{i}\dfrac{\mu_j}{\lambda_j}\Big) = \infty. \end{cases} \tag{7.9}$$

平均吸收时间为

$$\begin{cases} \infty & \text{若 } \sum\limits_{i=1}^{\infty}\rho_i = \infty, \\ \sum\limits_{i=1}^{\infty}\rho_i + \sum\limits_{r=1}^{m-1}\Big(\prod\limits_{k=1}^{r}\dfrac{\mu_k}{\lambda_k}\Big)\sum\limits_{j=r+1}^{\infty}\rho_j & \text{若 } \sum\limits_{i=1}^{\infty}\rho_i < \infty, \end{cases} \tag{7.10}$$

149 其中 $\rho_i = (\lambda_1\lambda_2\cdots\lambda_{i-1})/(\mu_1\mu_2\cdots\mu_i)$.

对于线性增长生灭过程的例子 ($\lambda_n = n\lambda, \mu_n = n\mu, \mu > \lambda$), 从状态 1 出发的平均吸收时间 ω_1 为

$$\sum_{i=1}^{\infty}\rho_i = \frac{1}{\mu}\sum_{i=1}^{\infty}\frac{1}{i}\Big(\frac{\lambda}{\mu}\Big)^{i-1} = \frac{1}{\lambda}\sum_{i=0}^{\infty}\int_0^{\lambda/\mu}\xi^i\mathrm{d}\xi = \frac{1}{\lambda}\int_0^{\lambda/\mu}\frac{1}{1-\xi}\mathrm{d}\xi = -\frac{1}{\lambda}\ln\Big(1-\frac{\lambda}{\mu}\Big). \tag{7.11}$$

4.8　有限状态连续时间马尔可夫链

一个连续时间马尔可夫链 $X_t(t > 0)$ 是状态为非负整数的马尔可夫过程. 我们通常假设转移概率是平稳的, 即

$$P_{ij}(t) = \Pr\{X_{t+s} = j | X_s = i\}. \tag{8.1}$$

在这一节, 我们仅仅考虑状态空间有限的情形, 记状态空间为 $\{0, 1, 2, \cdots, N\}$. 关于无限状态连续时间马尔可夫链的许多方面将在后面章节中讨论.

马尔可夫性表明 $P_{ij}(t)$ 满足:

(a) $P_{ij}(t) \geqslant 0$;

(b) $\sum\limits_{j=0}^{N} P_{ij}(t) = 1,\ i, j \in S$;

(c) $P_{ik}(s+t)=\sum_{j=0}^{N}P_{ij}(s)P_{jk}(t)$, $t,s\geqslant 0$ (切普曼 - 科尔莫戈罗夫方程);

此外, 我们假设

(d) $\lim_{t\to 0+}P_{ij}(t)=\begin{cases}1, & i=j,\\ 0, & i\neq j.\end{cases}$

如果 $\boldsymbol{P}(t)$ 表示矩阵 $\|P_{ij}(t)\|_{i,j=0}^{N}$, 则性质 (c) 用矩阵符号可简洁地表达为

$$\boldsymbol{P}(t+s)=\boldsymbol{P}(t)\boldsymbol{P}(s),\quad t,s\geqslant 0. \tag{8.2}$$

性质 (d) 说明 $\boldsymbol{P}(t)$ 在 $t=0$ 是连续的, 因为 (8.2) 蕴涵 $\boldsymbol{P}(0)=\boldsymbol{I}$(单位矩阵). 由 (8.2) 易推得 $\boldsymbol{P}(t)$ 对所有 $t>0$ 是连续的. 事实上, 在 (8.2) 中如果 $s=h>0$, 则由于 (d) 我们有

$$\lim_{h\to 0+}\boldsymbol{P}(t+h)=\boldsymbol{P}(t)\lim_{h\to 0+}\boldsymbol{P}(h)=\boldsymbol{P}(t)\boldsymbol{I}=\boldsymbol{P}(t). \tag{8.3}$$

150

另一方面, 对于 $t>0$ 和 $0<h<t$, 我们把 (8.2) 写为形式

$$\boldsymbol{P}(t)=\boldsymbol{P}(t-h)\boldsymbol{P}(h). \tag{8.4}$$

但当 h 足够小时, $\boldsymbol{P}(h)$ 近似于单位阵, 所以 $\boldsymbol{P}(h)^{-1}$[$\boldsymbol{P}(h)$ 的逆] 存在, 并且也接近于单位阵 $\boldsymbol{I}$. 因此,

$$\boldsymbol{P}(t)=\boldsymbol{P}(t)\lim_{h\to 0+}(\boldsymbol{P}(h))^{-1}=\lim_{h\to 0+}\boldsymbol{P}(t-h). \tag{8.5}$$

极限关系式 (8.3) 和 (8.5) 说明 $\boldsymbol{P}(t)$ 是连续的.

第 14 章的定理 1.1 和定理 1.2 证明了对于具有无限状态连续时间一般马尔可夫链, 下面极限存在:

$$\begin{aligned}&\lim_{h\to 0+}\frac{1-\boldsymbol{P}_{ii}(h)}{h}=q_i,\\&\lim_{h\to 0+}\frac{\boldsymbol{P}_{ij}(h)}{h}=q_{ij},\quad i\neq j,\end{aligned} \tag{8.6}$$

其中 $0\leqslant q_{ij}<\infty(i\neq j)$ 和 $0\leqslant q_i\leqslant\infty$, 即 $q_{ij}(i\neq j)$ 总是有限的, q_i 可以是无限的. $q_i=\infty$ 在有限状态连续时间的马尔可夫链中不可能出现, 事实上, 由关系式

$$1=\boldsymbol{P}_{ii}(h)+\sum_{j=0,j\neq i}^{N}\boldsymbol{P}_{ij}(h),$$

两边除以 h, 并令 h 递减趋于 0, 直接得到关系式

$$q_i=\sum_{j=0,j\neq i}^{N}q_{ij},$$

由此即知 q_i 确实是有限的.

假设 (8.6) 已被验证, 我们来导出一个 $\boldsymbol{P}_{ij}(t)$ 的显式表达式, 它使用下面无限小矩阵来表达:

$$\boldsymbol{A}=\left\|\begin{matrix} -q_0 & q_{01} & \cdots & q_{0N} \\ q_{10} & -q_1 & \cdots & q_{1N} \\ \vdots & & & \\ q_{N0} & q_{N1} & \cdots & -q_N \end{matrix}\right\|.$$

极限关系式 (8.6) 可用矩阵形式简单地表示为:

$$\lim_{h\to 0+}\frac{\boldsymbol{P}(h)-\boldsymbol{I}}{h}=\boldsymbol{A}. \tag{8.7}$$

借助于这个公式并回顾公式 (8.2), 我们有

151

$$\frac{\boldsymbol{P}(t+h)-\boldsymbol{P}(t)}{h}=\frac{\boldsymbol{P}(t)[\boldsymbol{P}(h)-\boldsymbol{I}]}{h}=\frac{\boldsymbol{P}(h)-\boldsymbol{I}}{h}\boldsymbol{P}(t). \tag{8.8}$$

右边极限存在, 从而导出矩阵微分方程式

$$\boldsymbol{P}'(t)=\boldsymbol{P}(t)\boldsymbol{A}=\boldsymbol{A}\boldsymbol{P}(t), \tag{8.9}$$

其中 $\boldsymbol{P}'(t)$ 表示元素为 $P'_{ij}(t)$ 的矩阵.

$\boldsymbol{P}'_{ij}(t)$ 的存在性显然是 (8.7) 和 (8.8) 的直接推论.

方程 (8.9) 在初始条件 $\boldsymbol{P}(0)=\boldsymbol{I}$ 下可按照常微分方程组理论 [1]解得

$$\boldsymbol{P}(t)=\mathrm{e}^{\boldsymbol{A}t}=\boldsymbol{I}+\sum_{n=1}^{\infty}\frac{\boldsymbol{A}^n t^n}{n!}. \tag{8.10}$$

在实际中若有可能, 我们可确定 $\boldsymbol{A}$ 的特征值 $\lambda_0,\lambda_1,\cdots,\lambda_N$ 和联系右特征向量 $\boldsymbol{u}^{(0)},\cdots,\boldsymbol{u}^{(N)}$ 的完备系 (见本书末的附录). 则我们有表示式:

$$\boldsymbol{P}(t)=\boldsymbol{U}\boldsymbol{\Lambda}(t)\boldsymbol{U}^{-1}, \tag{8.11}$$

其中矩阵 $\boldsymbol{U}$ 的列向量分别是 $\boldsymbol{u}^{(0)},\boldsymbol{u}^{(1)},\cdots,\boldsymbol{u}^{(N)}$, 并且 $\boldsymbol{\Lambda}(t)$ 是对角矩阵

$$\boldsymbol{\Lambda}(t)=\left\|\begin{matrix} \exp(\lambda_0 t) & 0 & \cdots & 0 \\ 0 & \exp(\lambda_1 t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \exp(\lambda_N t) \end{matrix}\right\|.$$

本章初等问题 7 和初等问题 13 中含有 (8.10) 或 (8.11) 的应用.

1. E. A. Coddington, and N.Levinson, *Theory of Ordinary Differential Equations*, Chapter 3. McGraw-Hill, New York, 1955.

初等问题

1. 设随机变量 X_1 和 X_2 是相互独立的且分别服从以 λ_1 和 λ_2 为参数的指数分布, 即

$$\Pr\{X_i > t\} = \exp\{-\lambda_i t\}, \quad t \geqslant 0, i = 1, 2.$$

152

令

$$N = \begin{cases} 1, & \text{若 } X_1 < X_2, \\ 2, & \text{若 } X_2 \leqslant X_1. \end{cases}$$
$$U = \min\{X_1, X_2\} = X_N,$$
$$V = \max\{X_1, X_2\},$$

且

$$W = V - U = |X_1 - X_2|.$$

试证明

(a) $\Pr\{N = 1\} = \lambda_1/(\lambda_1 + \lambda_2)$ 和 $\Pr\{N = 2\} = \lambda_2/(\lambda_1 + \lambda_2)$.

(b) 对于 $t \geqslant 0$ 有 $\Pr\{U > t\} = \exp\{-(\lambda_1 + \lambda_2)t\}$.

(c) N 和 U 是相互独立的随机变量.

(d) 对于 $t \geqslant 0$,

$$\Pr\{W > t | N = 1\} = \exp\{-\lambda_2 t\} \text{ 和}$$
$$\Pr\{W > t | N = 2\} = \exp\{-\lambda_1 t\}.$$

(e) U 和 $W = V - U$ 是相互独立的随机变量.

2. 设某装置在 k 次震动之后失灵. 若震动按参数为 λ 的泊松过程发生, 求该装置寿命 T 的密度函数.

解答:

$$f(t) = \begin{cases} \dfrac{\lambda^k t^{k-1} \mathrm{e}^{-\lambda t}}{\Gamma(k)}, & t > 0, \\ 0, & t \leqslant 0. \end{cases}$$

3. 设有参数为 λ 的泊松过程 $\{X(t), t \geqslant 0\}$. 若每次事件到达以概率 p 记录下来, 并且彼此间相互独立. 令 $\{Y(t), t \geqslant 0\}$ 是记录下来到达事件的过程, 试证明 $Y(t)$ 是参数为 λp 的泊松过程.

4. 设 $\{X(t), t \geqslant 0\}$ 和 $\{Y(t), t \geqslant 0\}$ 是参数分别为 λ_1 和 λ_2 的相互独立的泊松过程. 令 $Z_1(t) = X(t) + Y(t), Z_2(t) = X(t) - Y(t)$, $Z_3(t) = X(t) + k$, k 是正整数, 试确定上述过程中哪一个是泊松的, 并求其参数 λ.

5. 设到达一电报局的信息是按照平均每小时 3 个信息的泊松规律, 试求:

(a) 早晨八时至十二时没有信息到达的概率.

(b) 下午首次信息到达时刻的分布.

6. 设 $X(t)$ 是参数为 λ 的齐次泊松过程. 试确定 $X(t)$ 与 $X(t+\tau)(t > 0, \tau > 0)$ 之间的协方差, 即计算 $E[(X(t) - E[X(t)])(X(t+\tau) - E[X(t+\tau)])]$.

153

7. 连续时间马尔可夫链有两个状态, 分别记为 0 和 1. 在状态 0 的等候时间遵循参数为 $\lambda > 0$ 的指数分布. 在状态 1 的等候时间遵循参数为 $\mu > 0$ 的指数分布. 计算在时刻 0 处于状态 0 并在时刻 t 仍处于状态 0 的概率 $P_{00}(t)$.

解答:

$$P_{00}(t) = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu}\mathrm{e}^{-(\lambda+\mu)t}.$$

8. 在初等问题 7 中令 $\lambda = \mu$ 并定义 $N(t)$ 是系统在时间 $t \geqslant 0$ 内改变状态的次数, 试求 $N(t)$ 的概率分布.

解答:

$$\Pr\{N(t) = n\} \doteq \mathrm{e}^{-\lambda t}\frac{(\lambda t)^n}{n!}.$$

9. 设 $X(t)$ 为纯生连续时间马尔可夫链, 若

$$\Pr\{\text{在 } (t, t+h) \text{ 内发生一个事件} | X(t) = \text{奇数}\} = \lambda_1 h + o(h)$$
$$\Pr\{\text{在 } (t, t+h) \text{ 内发生一个事件} | X(t) = \text{偶数}\} = \lambda_2 h + o(h)$$

其中当 $h \downarrow 0$ 时 $o(h)/h \to 0$. 取 $X(0) = 0$. 试求下面概率:

$$P_1(t) = \Pr\{X(t) = \text{奇数}\}, \quad P_2(t) = \Pr\{X(t) = \text{偶数}\}.$$

提示: 导出微分方程

$$P_1'(t) = -\lambda_1 P_1(t) + \lambda_2 P_2(t),\ P_2'(t) = \lambda_1 P_1(t) - \lambda_2 P_2(t)$$

并解之.

解答:

$$P_1(t) = \frac{\lambda_2}{\lambda_1+\lambda_2}(1 - \exp\{-(\lambda_1+\lambda_2)t\});$$
$$P_2(t) = \frac{\lambda_1}{\lambda_1+\lambda_2} + \frac{\lambda_2}{\lambda_1+\lambda_2}\exp\{-(\lambda_1+\lambda_2)t\}.$$

10. 在初等问题 9 的条件下, 试求 $E[X(t)]$.

解答:

$$E[X(t)] = \frac{2\lambda_1\lambda_2}{\lambda_1+\lambda_2}t + \frac{(\lambda_1-\lambda_2)\lambda_2}{(\lambda_1+\lambda_2)^2}[\exp\{-(\lambda_1+\lambda_2)t\} - 1].$$

11. 假设 $g(t)$ 是一物品已知在时刻 t 前未损坏过而在时刻 t 损坏的条件比率, 即

$$\Pr\{\text{在时间 } (t, t+h) \text{ 内损坏} | \text{时刻 } t \text{ 前未损坏}\} = g(t)h + o(h), h \downarrow 0,$$

假设 $g(t)$ 是正的并且在 $(0, \infty)$ 连续. 试利用 $g(\cdot)$ 求 $F(t) = \Pr\{\text{在某个时刻 } \tau \text{ 损坏}, \tau < t\}$的表达式.

154

提示: 导出 $F(t)$ 的微分方程式.

解答:

$$F(t) = 1 - \exp\left[-\int_0^t g(\tau)\mathrm{d}\tau\right].$$

12. 考虑依时间变化的泊松过程, 即在时间区间 $(t, t+h)$ 内出现一个事件 E 与 E 之前事件出现次数独立, 并且其概率为 $\lambda(t)h+o(h)(h\to 0)$. (注意: λ 依赖于 t).

(a) 证明在时间区间 $[0,s]$ 内事件 E 不出现的概率为

$$\exp\left(-\int_0^s \lambda(\xi)\mathrm{d}\xi\right).$$

(b) 证明在时间区间 $[0,s]$ 内事件 E 出现 k 次的概率为

$$\frac{1}{k!}\left(\int_0^s \lambda(\xi)\mathrm{d}\xi\right)^k \exp\left(-\int_0^s \lambda(\xi)\mathrm{d}\xi\right).$$

13. 有两条横穿大西洋海底电缆, 每条每次可处理一电报信息. 每条电缆发生故障的时间均服从参数为 λ 的指数分布, 其修理所需时间均服从参数为 μ 的指数分布. 已知在时刻 0 的两条电缆工作正常, 若在时刻 t 有两个信息同时到达, 试求此时两条电缆均工作正常的概率.

提示: 这是具有三个状态连续时间马尔可夫链.

解答:

$$\frac{\mu^2}{(\lambda+\mu)^2}+\frac{\lambda^2\mathrm{e}^{-2(\lambda+\mu)t}}{(\lambda+\mu)^2}+\frac{2\lambda\mu}{(\lambda+\mu)^2}\mathrm{e}^{-(\lambda+\mu)t}.$$

14. 考虑具有参数 λ,μ 和 $a=0$ 的线性增长生灭过程. 设 $X(0)=1$. 求在第一个死亡时生物体数目的分布.

解答:

$$\Pr\{\text{在第一个死亡之前有 } k \text{ 个出生}\}=(\mu/(\mu+\lambda))(\lambda/(\lambda+\mu))^k.$$

15. 求线性增长生灭过程 (见 4.6 节例 1) 当 $\lambda<\mu$ 时平稳分布.

解答:

$$p_n=\left(\frac{\lambda}{\mu}\right)^n\frac{(a/\lambda)((a/\lambda)+1)\cdots((a/\lambda)+n-1)}{n!}\left(1-\frac{\lambda}{\mu}\right)^{a/\lambda}.$$

16. 一电话交换台有 m 条话路. 呼唤以参数为 λ 的泊松过程到达. 对于一呼唤到达, 若存在话路未被占用, 则此呼唤被接受, 否则丢失. 每次呼唤持续时间是一服从参数为 μ 的指数分布的随机变量, 并且它们彼此独立. 求被占用话路数目的平稳分布. 155

解答:

$$p_n=\frac{(\lambda/\mu)^n(1/n!)}{\sum_{k=0}^{m}(\lambda/\mu)^k(1/k!)},\quad n=0,1,2,\cdots,m.$$

17. 我们从时刻 0 开始观察一个放射性原子. 此原子衰变至时刻 $t(t>0)$ 停止放射, 其分布由下式确定,

$$F(\tau)=\left\{\begin{array}{ll}0, & \tau<0\\ 1-\mathrm{e}^{-\lambda\tau}, & \tau\geqslant 0\end{array}\right\}=\Pr\{t\leqslant\tau\}.$$

将此原子在时刻 t 时状态视为一随机变量

$$x_t = \begin{cases} 0, & \text{若原子在时刻 } t \text{ 是放射的}, \\ 1, & \text{若原子在时刻 } t \text{ 不是放射的}. \end{cases}$$

则 $\{x_t\}$ 构成一随机过程.

现假设在时刻 0 我们开始观察 N 个独立放射性原子, 按上述意义相应有随机变量 $x_t^i, i = 1, 2, \cdots, N$. 令 $X_t = \sum_{i=1}^{N} x_t^i$. 那么 $\{X_t\}$ 也是一随机过程. 试证明对于 $t \ll 1/\lambda$(意指 t 与 $1/\lambda$ 比较是可忽略地小) 和足够大的 N, $\{X_t\}$ 是非常近似于参数为 $\lambda N t$ 的泊松过程.

18. 设在问题 17 中不满足 $t \ll 1/\lambda$, 试问

(i) 此过程具有独立增量?　(ii) 它是否平稳?

(iii) 它是否有平稳转移概率?　(iv) 它是否马尔可夫过程?

解答: (i) 是, (ii) 不是, (iii) 是, (iv) 是

19. 本题是研究一移居生物群体灭亡时间.

令 $Z(t)$ 表示在时刻 t 群体总量. 若 $(Z(t); t \geqslant 0)$ 为一生灭过程, 其个体出生比率为 λ, 死亡比率为 μ. 这意味着每个个体彼此独立地在时间区间 $(t, t+\Delta t)$ 内以概率 $\lambda(\Delta t)$(近似地) 繁殖一个新个体, 而以概率 $\mu(\Delta t)$ 死亡.

我们想构造一种模型用来估计此种个体种群平均存活时间, 并希望用它来反映出种群数目均有最大上界的事实. 假设由于自然环境限制, 此群体最大限度的总量为 k 个个体. 由于所有个体都有可能死亡, 故当时间足够长时, 种群必将走向绝迹. 我们希望在这个模型中, 当群 [156] 体较小时具有指数增长的性质, 如上限 k. 超过最大限度则不能正常地繁殖. 这里有很多种方法使群体总量接近 k, 且平衡保持原状. 为简单, 我们取出生参数为

$$\lambda_i = \begin{cases} \lambda i, & \text{当 } i = 0, \cdots, k-1, \\ 0, & \text{当 } i \geqslant k, \end{cases}$$

死亡参数为 $\mu_i = \mu i (i = 0, 1, \cdots)$.

在上述模型之下, 经过一段时间, 此生物群体最后必定灭亡. 假设最初此群体仅有一个个体, 试计算其平均灭亡时间.

20. 设有一架机器失灵有两种形式. 第一种形式失灵发生在区间 $(t, t+h)$ 的概率为 $\lambda_1 h + o(h)$, 而第二种形式发生在区间 $(t, t+h)$ 的概率为 $\lambda_2 h + o(h)$. 当机器失灵时即进行修理, 修理持续时间服从指数分布, 其参数对应于第一种形式失灵的为 μ_1, 对应于第二种形式的为 μ_2. 试计算这架机器在时刻 t 正在工作的概率.

解答:

$$\boldsymbol{P}(t) = \mathrm{e}^{\boldsymbol{Q}t}, \quad \text{其中 } \boldsymbol{Q} = \left\| \begin{matrix} -(\lambda+\lambda_2) & \lambda_1 & \lambda_2 \\ \mu_1 & -\mu_1 & 0 \\ \mu_2 & 0 & -\mu_2 \end{matrix} \right\|.$$

21. 试比较 M|M|1 系统的两种排队规则, 一种是先来先服务, 另一种是后来先服务 (例如供服务用品是从一堆物品的顶部取的). 它们的排队长度, 等候时间以及服务忙碌期的分布是否不同?

解答：排队长度和服务忙碌期是相同的, 而等候时间的分布不同. (为什么?)

22. (有选择的排队过程) 设顾客依参数为 λ 的泊松过程到达柜台. 服务时间是独立同分布的, 其共同分布为参数 μ 的指数分布. 当顾客发现服务员 E 在服务时, 以概率 $p(0 < p < 1)$ 加入队伍. 试用生灭过程描述此模型.

解答：$\lambda_n = \lambda p$; $\mu_n = \mu$.

23. 考虑规则为后来先服务的 $M|M|1$ 系统. 令 $X(t)$ 表示在时刻 t 的排队长度, 试证明 $\{X(t); t \geqslant 0\}$ 是一生灭过程并确定其参数.

解答：$\lambda_n = \lambda$, $\mu_n = \mu$.

24. 在条件 $X(0) = N = 1$ 之下, 试确定 Yule 过程的均值和方差.

解答：$E[X(t)] = \mathrm{e}^{\lambda t}$, $\mathrm{Var}[X(t)] = \mathrm{e}^{2\lambda t}(1 - \mathrm{e}^{-\lambda t})$.

157

问　题

1. 设 $\{X(t), t \geqslant 0\}$ 和 $\{Y(t), t \geqslant 0\}$ 分别是参数为 λ_1 和 λ_2 的两个相互独立泊松过程. 令

$$Z(t) = X(t) - Y(t), \quad t \geqslant 0.$$

这是状态空间为全体整数的随机过程. 令

$$P_n(t) = \Pr\{Z(t) = n\}, \quad n = 0, \pm 1, \pm, 2, \cdots.$$

试证明公式

$$\sum_{n=-\infty}^{\infty} P_n(t) z^n = \exp(-(\lambda_1 + \lambda_2)t) \exp(\lambda_1 z t + (\lambda_2/z)t), \quad |z| \neq 0,$$

并计算 $E[Z(t)]$ 和 $E[Z(t)^2]$.

答案：$E[Z(t)]^2 = (\lambda_1 + \lambda_2)t + (\lambda_1 - \lambda_2)^2 t^2$.

2. 考虑两个独立泊松过程 $X(t)$ 和 $Y(t)$, 其中 $E[X(t)] = \lambda t, E[Y(t)] = \mu t$. 对于过程 $X(t)$, 有两个事件发生在 T 和 $T' > T$, 使得当 $T \leqslant t < T'$ 时有 $X(t) = X(T)$. 并且 $X(T') = X(T) + 1$. 定义 $N = Y(T') - Y(T) =$ 发生在区间 (T, T') 内过程 $Y(t)$ 的事件数. 试求 N 的分布.

答案：$\Pr\{N = m\} = \dfrac{\lambda}{\lambda + \mu}\left(\dfrac{\mu}{\lambda + \mu}\right)^m, \quad m = 0, 1, 2, \cdots.$

3. 考虑一纯灭过程, 其中 $\mu_n = n\mu$, $n = 1, 2, \cdots$, 即当 $j > k, k, t > 0$ 时有 $\mathrm{P}\{X(t+h) = j | X(t) = k\} = 0$. 设初始群体总量为 i. 试求 $P_n(t) = \mathrm{P}\{X(t) = n\}$, $E[X(t)]$ 和 $\mathrm{Var}(X(t))$.

答案:

$$P_n(t) = \binom{i}{n} \mathrm{e}^{-n\mu t}(1-\mathrm{e}^{-\mu t})^{(i-n)},$$

$$E[X(t)] = i\mathrm{e}^{-\mu t},$$

$$\mathrm{Var}X(t) = i\mathrm{e}^{-\mu t}(1-\mathrm{e}^{-\mu t}).$$

4. 考虑参数为 $\beta = 1$, 初始状态为 $N = 1$ 的 Yule 过程. 设第一个个体将死亡, 当已知此个体在时刻 t 活着而在区间 $(t, t+h)$ 死亡的概率为 $\mu h + o(h)$. 试求最初单一个体及此个体死亡时刻其后代所产生后代总数的分布.

158

答案: 由指定的一个个体和此个体死亡时刻其后代所产生后代总数为 n 的概率为

$$\int_0^{\infty} \mathrm{e}^{-\beta t}(1-\mathrm{e}^{-\beta t})^n \mu \mathrm{e}^{-\mu t}\mathrm{d}t = \frac{\mu}{\beta}\frac{\Gamma((\mu/\beta)+1)\Gamma(n+1)}{\Gamma(n+(\mu/\beta)+2)}.$$

5. 设 $(X(t), Y(t))$ 是二维随机过程, 其中 $X(t)$ 是参数为 λ_1 的泊松过程, $Y(t)$ 是与 $X(t)$ 独立的参数为 λ_2 的泊松过程. 已知过程在 $t=0$ 时处于状态 (x_0, y_0), 此处 $x_0 + y_0 < z$, 试求状态与直线 $x + y = z$ 相交于点 (x, y) 的概率.

答案:

$$\begin{cases} \dbinom{z-x_0-y_0}{x-x_0}\left(\dfrac{\lambda_1}{\lambda_1+\lambda_2}\right)^{x-x_0}\left(\dfrac{\lambda_2}{\lambda_1+\lambda_2}\right)^{y-y_0}, & x \geqslant x_0, y \geqslant y_0, \\ 0, & \text{其他}. \end{cases}$$

6. 考虑参数为 λ 的泊松过程, 设 T 为观察到第一个事件所需的时间, 令 $N(T/k)$ 表示紧接着的 T/k 单位时间里事件发生的次数. 试求 $N(T/k)T$ 的前两阶矩.

答案:

$$E\left\{N\left(\frac{T}{k}\right)T\right\} = \frac{2}{\lambda k}; \quad E\left\{\left(N\left(\frac{T}{k}\right)T\right)^2\right\} = \frac{6}{\lambda^2 k} + \frac{24}{\lambda^2 k^2}.$$

7. 考虑 n 个独立对象 (诸如灯泡), 其损坏时间 (即寿命) 遵循指数分布, 其密度函数为 $f(x,\theta) = \theta^{-1}\exp(-x/\theta), x > 0$; 当 $x \leqslant 0$ 时, 密度函数为 0(θ 是正数). 按损坏的顺序观察 r 个对象, 令

$$X_{1,n} \leqslant X_{2,n} \leqslant \cdots \leqslant X_{r,n}$$

表示首先损坏的 r 个对象的寿命. 试确定 $X_{i,n}, i = 1, 2, \cdots, r$ 的联合密度函数.

答案:

$$f(x_1, x_2, \cdots, x_r) = r!\binom{n}{r}\frac{1}{\theta^r}\exp\left(-\frac{x_1 + x_2 + \cdots + x_{r-1} + (n-r+1)x_r}{\theta}\right).$$

8. 在上面问题中定义 $Y_{1,n} = X_{1,n}$ 和

$$Y_{i,n} = X_{i,n} - X_{i-1,n}, \quad 2 \leqslant i \leqslant r.$$

试证明 $Y_{i,n}$ 是相互独立的并求其各自的分布函数.

答案：

$$\Pr\{Y_{i,n} \leqslant y\} = 1 - \exp\left(-\frac{n-i+1}{\theta}y\right).$$

159

9. 设有一参数为 λ 的泊松过程. 若已知在时间 t 内发生了 n 个事件, 试求出现 r 个 $(r < n)$ 事件的时间密度函数.

答案：

$$p(x) = \begin{cases} \dfrac{n!}{(r-1)!(n-r)!}\dfrac{x^{r-1}}{t^r}\left(1-\dfrac{x}{t}\right)^{n-r}, & 0 < x < t, \\ 0, & \text{其他}. \end{cases}$$

10. 设 $\mathcal{M}$ 是连续时间生灭过程, 其中 $\lambda_n = \lambda > 0$, $n \geqslant 0$, $\mu_0 = 0$, $\mu_n > 0, n \geqslant 1$. 令 $\pi = \sum_n \pi_n < \infty$, 此处 $\pi_n = \lambda^n/(\mu_1\mu_2\cdots\mu_n)$, π_i/π 是过程的平稳分布. 假定初始状态是一随机变量, 其分布就是该过程的平稳分布. 试证明在 $[0,t]$ 内死亡的数量具有参数为 λt 的泊松分布.

提示：设 $a_k(t)$ 是至时刻 t 为止死亡数为 k 的概率. 导出微分方程式

$$a_k'(t) = -\lambda a_k(t) + \lambda a_{k-1}(t), \quad k = 1, 2, \cdots.$$

11. 下面定义为二维情况下多变量泊松过程的定义. 设 $(X(t), Y(t))$ 由 $X(t) = \alpha(t) + \gamma(t), Y(t) = \beta(t) + \gamma(t)$ 所确定, 其中 $\alpha(t)$, $\beta(t)$ 和 $\gamma(t)$ 是参数分别为 λ_1, λ_2 和 λ_3 的彼此独立的泊松过程, 试求 $(X(t), Y(t))$ 分布的概率母函数.

答案：

$$\begin{aligned} &\sum \Pr\{X(t) = i, Y(t) = j\}x^i y^j \\ &= \exp\{t(\lambda_1 x + \lambda_2 y + \lambda_3 xy - \lambda_1 - \lambda_2 - \lambda_3)\}. \end{aligned}$$

12. 设有一出生比率为 λ 及初始群体总量为 1 的 Yule 过程 $\{N_t, t \geqslant 0\}$. $N_t(x)$ 表示群体在时刻 t 年龄不超过 x 的成员数. 试求 $N_t(x)$ 的分布函数.

提示：对 N_{t-x} 取值进行假设.

答案：

$$\Pr\{N_t(x) = k\} = \frac{e^{-\lambda t}(1 - e^{-\lambda x})^k}{[1 - e^{-\lambda x} + e^{-\lambda t}]^{k+1}}.$$

13. 设 $\{X_i(t); t \geqslant 0\}$, $i = 1, 2$, 是具有相同参数 λ 的相互独立的 Yule 过程, 且 $X_i(0) = n_i$, $i = 1, 2$. 试确定当已知 $X_1(t) + X_2(t) = N(N \geqslant n_1 + n_2)$ 时, $X_1(t)$ 的条件分布.

答案：

$$\Pr\{X_1(t) = k | X_1(t) + X_2(t) = N\} = \frac{\dbinom{k-1}{n_1-1}\dbinom{N-k-1}{n_2-1}}{\dbinom{N-1}{n_1+n_2-1}}, \quad k \geqslant n_1$$

160

14. 继续问题 13. 试证明极限分布关系式：

$$\lim_{t\to\infty} \Pr\left\{\frac{X_1(t)}{X_1(t) + X_2(t)} \leqslant x\right\} = \frac{(n_1+n_2-1)!}{(n_1-1)!(n_2-1)!}\int_0^x y^{n_1-1}(1-y)^{n_2-1}dy.$$

提示：令 $N \to \infty$ 和 $k \to \infty$ 且 $k/N \to y(0 < y < 1)$. 利用斯特林公式建立渐近关系式

$$\lim_{\frac{k}{N}\to y,\ N\to\infty,\ k\to\infty} \frac{\binom{k-1}{n_1-1}\binom{N-k-1}{n_2-1}}{\binom{N-1}{n_1+n_2-1}} = \frac{(n_1+n_2-1)!}{(n_1-1)!(n_2-1)!} y^{n_1-1}(1-y)^{n_2-1},$$

然后证明

$$\begin{aligned}&\lim_{t\to\infty} \Pr\left\{y \leqslant \frac{X_1(t)}{X_1(t)+X_2(t)} \leqslant y+h\right\} \\ &= \frac{(n_1+n_2-1)!}{(n_1-1)!(n_2-1)!} h y^{n_1-1}(1-y)^{n_2-1} + o(h).\end{aligned}$$

15. 设一个系统是由 N 个相同部件构成的. 每个部件彼此独立地运转随机长度时间, 直到损坏. 若损坏时间分布是参数为 λ 的指数分布. 当一个部件损坏时需加修理, 修理时间是随机的, 遵循参数为 μ 的指数分布. 如果在时刻 t. 正好有 n 个部件正在修理, 就说系统在时刻 t 处于状态 n. 此过程是生灭过程, 试确定其无穷小参数.

16. 在问题 15 中, 设最初 N 个部件都在运转. 试求有两个部件不运转的首次时间分布 $F(t)$.

答案：$F(t)$ 的拉普拉斯变换是

$$\varphi(s) = \frac{N(N-1)\lambda^2}{s^2 + s[(2N-1)\lambda+\mu] + N(N-1)\lambda^2}.$$

在情况 $\lambda = \mu$ 下,

$$1 - F(t) = \frac{\sqrt{N}(N-1)}{2}\{\exp[(-N+\sqrt{N})\lambda t] - \exp[(-N-\sqrt{N})\lambda t]\}.$$

17. 考虑下述 Ehrenfest 模型 (见 2.2 节中的 B) 的连续变型. 我们有 $2N$ 个球, 标有号码 $1, 2, 3, \cdots, 2N$. 在时刻 0 每个球等可能被放置于两个罐中之一. 随后, 这些球按下面规则从一个罐到另一个罐进行随机替换. 一个球在时间区间 $(t, t+h)$ 内以概率 $h/2 + o(h)$ 改变罐子, 以概率 $1 - (h/2) + o(h)$ 留在原来罐子. 这些替换在不相交时间区间内是相互独立的. 令 $X(t)$ 表示在时刻 t 罐子 I 里球的数目. 置

161

$$P_{jk}(t) = \Pr\{X(t) = k | X(0) = j\}, \quad j, k = 0, 1, \cdots, 2N.$$

试证明公式

$$g(t,s) = \sum_{k=0}^{2N} P_{ik}(t) s^k = 2^{-2N}[1 - \mathrm{e}^{-t} + (1+\mathrm{e}^{-t})s]^j [1 + \mathrm{e}^{-t} + (1-\mathrm{e}^{-t})s]^{2N-j}.$$

提示：定义随机变量

$$X_i(t) = \begin{cases} 1, & \text{若第 } i \text{ 个球于时刻 } t \text{ 在罐子 } I,\ i = 1, 2, \cdots, 2N, \\ 0, & \text{其他,} \end{cases}$$

则

$$X(t) = \sum_{i=1}^{2N} X_i(t).$$

试证明

$$\Pr\{X_i(t) = X_i(0)\} = \frac{1+e^{-t}}{2}, \quad i = 1, 2, \cdots, 2N.$$

18. 设有一线性增长生灭过程 $X(t)$, 其中 $\lambda = \mu$(见 4.6 节例 1), 试证明

$$u(t) = \Pr\{X(t) = 0 | X(0) = 1\}$$

满足积分方程

$$u(t) = \frac{1}{2}\int_0^t 2\lambda e^{-2\lambda\tau} d\tau + \frac{1}{2}\int_0^t 2\lambda e^{-2\lambda\tau}[u(t-\tau)]^2 d\tau.$$

提示: 等候第一个事件 (生或灭) 的时间遵循参数为 2λ 的指数分布.

19. (续上) 证明 $u(t)$ 满足 Ricatti 微分方程式

$$u'(t) + 2\lambda u(t) = \lambda + \lambda u^2(t), \quad u(0) = 0.$$

20. (续上) 求 $u(t)$

答案:

$$u(t) = \frac{\lambda t}{1+\lambda t}.$$

21. (续上) 求 $\Pr\{X(t) = 0 | X(0) = 1, X(T) = 0\}$, 其中 $0 < t < T$.

22. 考虑具有无限小参数为 λ_n, μ_n 的生灭过程. 试证明从状态 0 出发到达状态 $r+1$ 平均的时间为

$$\sum_{n=0}^{r} \frac{1}{\lambda_n \pi_n} \sum_{k=0}^{n} \pi_k,$$

其中 π_n 的定义见等式 (4.5).

162

提示: 设 T_n^* 表示从状态 n 出发第一次到达状态 $n+1$ 所需时间, 试导出关于 $E(T_n^*)$ 的递推关系式.

23. 下面的问题源出于分子生物学. 细菌表面假设由几处组成, 在每处, 若成分适当异体分子便可附着其上. 这样成分的分子称为可接受的. 我们考虑某一处, 假设分子到达此处按照参数为 μ 的泊松过程. 这些分子中有比例为 β 是可接受的. 不可接受的分子停留在该处的时间服从参数为 λ 的指数分布. 当它停留在该处时, 其他分子皆不附着在那里, 而可接受的分子能 "固定" 在此处, 使得任何其他分子不再能附着在那里. 试问在本题中该处至时刻 t 为止未被分子 "固定" 的概率.

提示: 将问题化归为三个状态连续时间马尔可夫链来研究.

答案:

$$\frac{\beta\mu}{s_2 - s_1}\left[\left(1+\frac{\lambda}{s_1}\right)e^{s_1 t} - \left(1+\frac{\lambda}{s_2}\right)e^{s_2 t}\right]$$

其中 s_1, s_2 是 $s^2 + s(\lambda+\mu) + \mu\beta\lambda = 0$ 的根.

24. 考虑一具有无限多服务的排队系统, 服务时间按照参数为 μ 的指数分布. 设顾客成批地到达, 其间隔时间按照参数为 λ 的指数分布, 并设各批顾客数服从参数为 $\rho(0<\rho<1)$ 的几何分布, 即 Pr{一批顾客数为 k}$=\rho^{k-1}(1-\rho)(k=1,2,\cdots)$. 试用连续时间马尔可夫过程描述此过程, 并具体确定其无限小矩阵.

25. 设有一处于平稳状态的 M|M|1 排队过程, 试证明其离开时间的间距分布与到达时间的间距分布是相同的 (参见问题 10).

26. 设 $\{X_i(t);t\geqslant 0\},i=1,2$, 是独立泊松过程, 其参数分别为 λ_1 和 λ_2. 若 $X_1(0)=m, X_2(0)=N-1$, 且 $m<N$.

(a) 试求过程 X_2 先于 X_1 到达 N 的概率.

(b) 若 $X_2(0)=n$, 其中 $n<N$, 试解同样问题.

答案: (b):

$$\sum_{r=0}^{N-m-1}\binom{N-n+r-1}{r}p^r q^{N-n},\quad p=\frac{\lambda_1}{\lambda_1+\lambda_2}.$$

163

27. 下面两个生灭过程 (见 4.4 节) 可看作有选择的排队过程模型.

(a) 考虑一生灭过程, 其参数为

$$\begin{aligned}&\lambda_n=\lambda q^n,\quad 0<q<1,\quad \lambda>0(n=0,1,2,\cdots),\\&\mu_n=\mu,\qquad\qquad\qquad\ \ \mu>0,\\&\mu_0=0.\end{aligned}$$

(b) 参数为

$$\begin{aligned}&\lambda_n=\frac{\lambda}{n+1},\mu_n=\mu(n=1,2,\cdots),\\&\mu_0=0.\end{aligned}$$

试确定每种情况下平稳分布.

答案: (a) 对于 $m\geqslant 1$, $p_m=p_0(\lambda/\mu)^m q^{m(m-1)/2}$

(b) 对于 $m\geqslant 1$, $p_m=p_0(\lambda/\mu)^m(1/m!)$, 其中 $p_0=e^{-\lambda/\mu}$.

28. 试证明 $M|M|s$ 排队系统的平稳排队长度分布 $\{p_n,n=0,1,2,\cdots\}$ 由下式给出:

$$p_0=\left\{\frac{(s\rho)^s}{s!(1-\rho)}+\sum_{i=0}^{s-1}\frac{(s\rho)^i}{i!}\right\}^{-1}$$

$$p_n=\begin{cases}p_0\dfrac{(s\rho)^n}{n!}, & 1\leqslant n\leqslant s,\\ p_0\rho^n\dfrac{s^s}{s!}, & s<n<\infty,\end{cases}$$

其中 $\rho=\lambda/s\mu<1$. 设 $Q=\max(n-s,0)(n=0,1,2,\cdots)$ 表示不包括正在受服务成员的排队长度, 试证明

(i) $$\gamma=\Pr\{Q=0\}=\frac{\displaystyle\sum_{i=0}^{s}(s\rho)^i/i!}{\displaystyle\sum_{i=0}^{s}[(s\rho)^i/i!]+[(s\rho)^s\rho/s!(1-\rho)]},$$

(ii) $E(Q) = (1-\gamma)/(1-\rho)$.

29. 设一系统由 N 部机器组成. 在任何时候至多 $M(\leqslant N)$ 部机器运转, 其余留作备用. 当一机器运转时就一直运转下去直到失灵, 失灵时间服从参数为 μ 的指数分布.

机器失灵时需加修理. 在任何时候至多 R 部机器处于修理之中. 设修理时间服从参数为 λ 的指数分布. 这样一部机器可能处于四种状态: (i) 运转, (ii) 正常但不运转即备用, (iii) 正在修理, (iv) 等候修理. 系统中共有 N 部机器. 至多有 M 部正在运转, 且至多有 R 部正在修理.

令 $X(t)$ 表示在时刻 t 处于正常状态 (运转或备用) 机器数. 那么, (我们假定) 运转机器数为 $\min\{X(t), M\}$ 及备用机器数为 $\max\{0, X(t)-M\}$. 而失灵机器数为 $Y(t) = N - X(t)$, 其中正在修理的为 $\min\{Y(t), R\}$, 等候修理的为 $\max\{0, Y(t)-R\}$. 这些公式使我们一旦知道 $X(t)$ 即可确定机器处于任何状态的数目, $X(t)$ 是一生灭过程.

164

(a) 试确定其生和灭的参数, λ_i 和 μ_i, $i = 0, \cdots, N$.

(b) 分别就情况 (a) $R = M = N$ 和 (b) $R = 1, M = N$, 求 $X(t) = j$ 的平稳概率 π_j.

30. (续上)(a) 在 $R < M = N$ 的情况下求其平稳分布. (b) 设 $X_{N,R}(\infty)$ 表示具有 (a) 中平稳分布的随机变量. 令 $N \to \infty, R \to \infty$, 并使得 $N/R \to 1+\alpha$, 其中 α 是固定的. 试确定其规范化常数 a_N, b_N 和极限分布 $\Phi(x)$, 使得

$$\lim_{N\to\infty} \Pr\left\{\frac{X_{N,R}(\infty) - a_N}{b_N} \leqslant x\right\} = \Phi(x).$$

(存在两种情况: (i) $\alpha > \lambda/\mu$ 和 (ii)$\alpha < \lambda/\mu$).

31. 考虑一无穷小参数为 $\lambda_n = \lambda n^2$ 的纯生过程, 此处, $\lambda > 0$. 已知在时刻 0 有一粒子, 试求

$$P_\infty(t) = 1 - \sum_{k=1}^{\infty} P_k(t).$$

32. 设 $X(t)$ 为一 Yule 过程, $X(0) = N$, 其出生比率为 β. 试证明

$$\Pr\{X(t) \geqslant n | X(0) = N\} = \sum_{k=n-N}^{n-1} \binom{n-1}{k} p^k q^{n-1-k},$$

其中 $q = 1 - p = \mathrm{e}^{-\beta t}$.

附　记

泊松过程和生灭过程在诸如排队和存储模型, 群体增长, 工程系统等理论和应用中起着非常重要的作用. 关于泊松过程以及与其有关过程的初步讨论可在任何一本随机过程教科书中找到.

有关排队理论的文献相当丰富. 总结这方面理论及其应用的一本优秀专著是 Cox 和 Smith[1].

读者也可参阅高级著作 Takács[2] 和 Riordan[3].

165

关于排队理论的概要可参看 Saaty[4]. 这本参考书还包括了广泛文献目录.

关于堆积理论和电话系统问题的应用可见书 Syski[5].

某些专门的排队理论数学问题论述请见 Beneš 的专著 [6].

参考书目

[1] D. R. Cox and W. L. Smith, *Queues*. Methuen, London, 1961.

[2] L. Takács, *Introduction to the Theory of Queues*. Oxford Univ. Press, London and New York, 1962.

[3] J. Riordan, *Stochastic Service Systems*. Wiley, New York, 1962.

[4] T. L. Saaty, *Elements of Queueing Theory with Applications*. McGraw-Hill, New York, 1961.

[5] R. Syski, *Congestion Theory*. Oliver and Boyd, New York, 1960.

[6] V. E. Beneš, *General Stochastic Processes in the Theory of Queues*. Addison-Wesley, Reading, Massachusetts, 1963.

166

第5章　更新过程

更新理论始于研究随时间变化而不断更新的随机系统, 它在每次更新之后依统计意义又重复原来变化过程. 目前更新理论着眼于研究一列表示更新时间区间的一般独立同分布且非负随机变量序列. 它的结果在理论和实际概率模型中有着广泛的应用.

本章的前 6 节是必不可少的, 它包含在每一本导引性教程中. 5.7 节和 5.8 节不太困难, 如果有时间读它, 可使更新理论的基本知识得以完满. 5.9 节提供了一个新近有趣论题的轮廓, 初次阅读可从略.

5.1　更新过程的定义和有关概念

更新 (计数) 过程 $\{N(t), t \geqslant 0\}$ 是非负整数值随机过程, 它是在时间区间 $(0, t]$ 内重复试验某事件陆续发生的次数, 而这些相继 "事件" 发生的间隔时间是一系列独立同分布正的随机变量. 记这一系列事件发生的间隔时间为 $\{X_k\}_{k=1}^{\infty}$(它通常表示陆续替换到装置里的元件寿命), 即 X_i 表示第 $i-1$ 个事件发生至第 i 个事件发生的间隔时间, 有时称为第 i 个更新间距. 我们记

$$F(x) = \Pr\{X_k \leqslant x\}, \quad k = 1, 2, 3, \cdots,$$

作为 $\{X_k\}$ 的共同概率分布. 更新过程的一个基本规定是 $F(0) = 0$, 这意味着 X_k 是正随机变量. 我们记

$$S_n = X_1 + X_2 + \cdots + X_n, \quad n \geqslant 1 (约定 S_0 = 0). \tag{1.1}$$

它是第 n 个事件出现之前的**等候时间**. 因此 (见图 5-1)

$$N(t) = 使\ 0 < S_n \leqslant t\ 成立的指标\ n\ 的数目. \tag{1.2}$$

通常, 计数过程 $\{N(t), t \geqslant 0\}$ 与部分和过程 $\{S_n, n \geqslant 0\}$ 都称为 "更新过程". 物理上原始的更新模型是一系列灯泡替换. 一个灯泡在时刻 0 被安装使用, 在时刻 X_1 烧坏, 然后换上一个新的. 第二个灯泡在时刻 $X_1 + X_2$ 烧坏, 然后换上第三个灯泡. 一般来说, 第 n 个灯泡在时刻 $\sum_{i=1}^{n} X_i$ 烧坏, 然后马上替换新的, 等等. 我们自然假定这些灯泡的寿命是统计独立的, 并且从概率意义上具有相同的特征, 即

$$\Pr\{X_k \leqslant x\} = F(x).$$

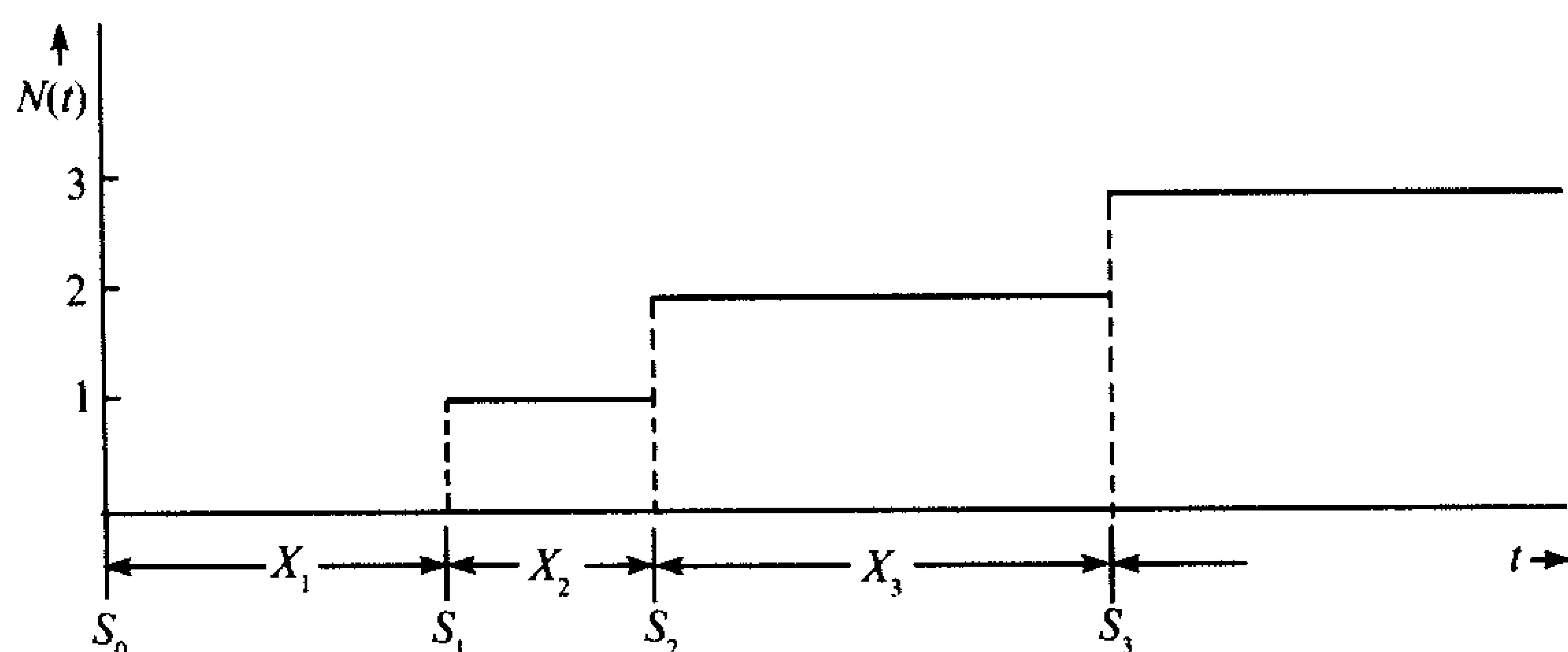


图 5-1 独立同分布随机变量序列 $\{X_n\}$ 与更新计数过程 $N(t)$ 之间的关系

显然, 在这个过程中, $N(t)$ 记录下至时刻 t 为止更新 (灯泡) 的数目.

更新理论主要目的是依据事件出现的间隔时间分布 F 导出与 $\{N(t)\}$ 和 $\{S_n\}$ 相联系的某些随机变量的性质. 例如很有意义的是计算在时间区间 $(0,t]$ 内更新数目的期望值:

$$E[N(t)] = M(t) \quad \text{称为}\textbf{更新函数}.$$

为此, 一些有关的关系式需要介绍. 原则上, $S_n = X_1 + \cdots + X_n$ 的概率分布.

168

$$\Pr\{S_n \leqslant x\} = F_n(x),$$

可利用卷积公式来计算, 其中 $F_1(x) = F(x)$ 假定是已知的, 我们有

$$F_n(x) = \int_0^\infty F_{n-1}(x-y)\mathrm{d}F(y) = \int_0^x F_{n-1}(x-y)\mathrm{d}F(y).^1$$

前面我们曾强调 (1.2) 作为过程 $\{S_n\}$ 和 $\{N(t)\}$ 的联系桥梁. 关系式 (1.2) 还可表示成形式:

$$N(t) \geqslant k \quad \text{当且仅当}\ S_k \leqslant t. \tag{1.3}$$

由此立刻推得

$$\begin{aligned}\Pr\{N(t) \geqslant k\} &= \Pr\{S_k \leqslant t\} \\ &= F_k(t), \quad t \geqslant 0, \quad k = 1,2,\cdots,\end{aligned} \tag{1.4}$$

因此

$$\begin{aligned}\Pr\{N(t) = k\} &= \Pr\{N(t) \geqslant k\} - \Pr\{N(t) \geqslant k+1\} \\ &= F_k(t) - F_{k+1}(t), \quad t \geqslant 0, \quad k = 1,2,\cdots.\end{aligned} \tag{1.5}$$

1. 这里和后面, 关于积分区间我们约定只包含右端点而不包括左端点, 即

$$\int_a^b h(x)\mathrm{d}G(x) = \int_{a+}^{b+} h(x)\mathrm{d}G(x).$$

由定义并注意到 (1.4)、(1.5), 我们得到

$$M(t)=E[N(t)]=\sum_{k=1}^{\infty}k\Pr\{N(t)=k\}$$

$$=\sum_{k=1}^{\infty}\Pr\{N(t)\geqslant k\}=\sum_{k=1}^{\infty}\Pr\{S_k\leqslant t\}=\sum_{k=1}^{\infty}F_k(t).$$

(第三个等号通过分部求和推得, 也可参考第 1 章初等问题 1). 级数

$$M(t)=\sum_{k=1}^{\infty}F_k(t) \tag{1.6}$$

的收敛性将在 5.4 节中证明.

问题 3、问题 16 和问题 17 涉及 $N(t)$ 的方差和高阶矩.

尚有一些其他有趣的随机变量. 下面列出三个: **剩余寿命**(也称为剩余随机变量), **现龄**(也称为现龄随机变量) 和**全寿命**. 其定义如下: 169

$$\gamma_t=S_{N(t)+1}-t \quad (\text{剩余或残留寿命})$$
$$\delta_t=t-S_{N(t)} \quad (\text{现龄或现龄随机变量})$$
$$\beta_t=\gamma_t+\delta_t \quad (\text{全寿命})$$

这些随机变量的图形表示见图 5-2. 关于这些随机变量在更新过程理论和实际应用中的基本意义, 本章后面将予以充分说明.

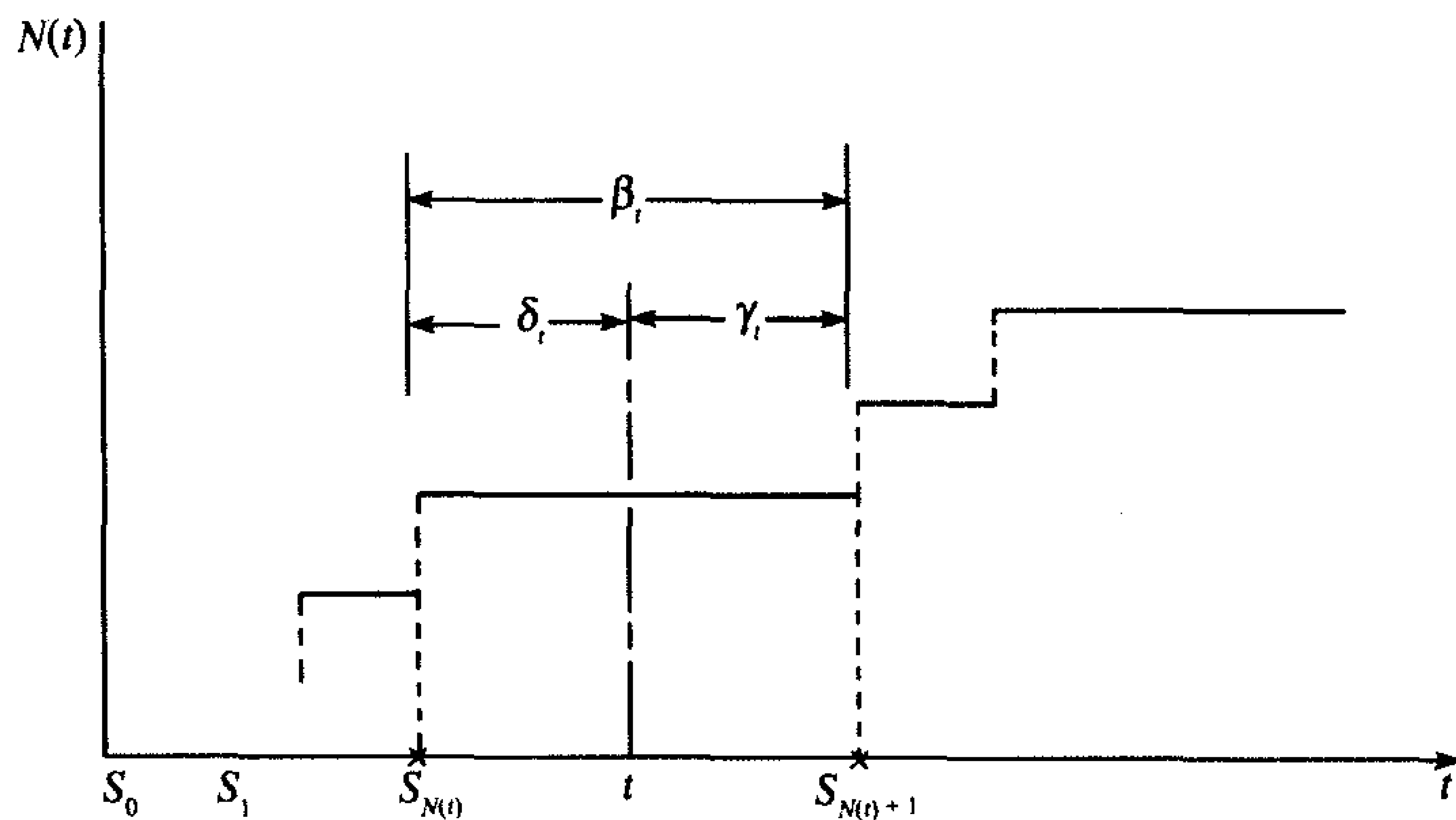


图 5-2 剩余寿命 γ_t, 现龄 δ_t, 全寿命 β_t

5.2 更新过程的一些例子

下面例子表明, 引起更新过程的范围是十分广泛的, 原因是多种多样的. 其中几个例子在后面几节将会被深入研究.

(a) 泊松过程

参数为 λ 的泊松过程 $\{N(t), t \geqslant 0\}$ 是更新计数过程, 其事件出现间隔时间服从指数分布 (见第 4 章定理 2.1)

$$F(x) = 1 - \mathrm{e}^{-\lambda x}, \quad x \geqslant 0,$$

170 这个特殊的更新过程具有许多特别性质, 将在 5.3 节中对此作详细介绍.

(b) 计数过程

相继输入到记录装置 (计数器) 的电脉冲或信号的间隔时间常常被认为构成一个更新过程. 大部分物理上可实现的计数器在记录一个脉冲后将闭锁一段时间, 在这段时间内不记录到达的其他脉冲. 计数器仅当处于开放 (即非闭锁) 状态时才记录脉冲. 在适当假设之下, 记录脉冲的时间序列构成一个更新过程. 但应强调指出, 记录脉冲的更新过程是二级更新过程, 它是由所有到达的脉冲 (其中包括未被记录的脉冲) 构成的原来更新过程导出的. 5.3 节与 5.7 节将详细说明两种闭锁类型的计数器理论, 即所谓 I 型和 II 型计数器.

(c) 交通流

通常假定在无限长的单行公路上相继行驶汽车之间距离构成一个更新过程. 相继行驶的汽车经过一个固定站的间隔时间也构成一个更新过程.

(d) 与排队联系的更新过程

在单一服务的排队过程中, 有很多更新过程嵌入在其中. 我们举两个例子:

(i) 如果顾客到达时间构成一个更新过程, 那么一系列服务期 (即忙碌时间) 的起始时刻构成第二级更新过程.

(ii) 若输入过程 (顾客到达的模型) 是泊松的, 服务期结束时刻序列也是一个更新过程.

(e) 存储系统

在分析大部分的存储模型中, 通常假定需求模型是一更新过程. 大部分标准存储策略可归结为更新过程. 例如, 进货时间是一更新过程 (见 5.8 节).

(f) 与独立随机变量和相联系的更新过程

(i) 设 $\xi_1, \xi_2, \xi_3, \cdots$ 是一列实值 (不一定是正的)i.i.d.(独立同分布) 随机变量. 假设 $E(\xi_i) \geqslant 0$. 考虑部分和 $U_0 = 0, U_n = \xi_1 + \xi_2 + \cdots + \xi_n$, $n \geqslant 1$, 其值域为实直线.
171

定义

$$S_1 = \inf\{n : U_n > 0\} \quad 和 \quad S_k = \inf\{n : U_n > U_{S_{k-1}}\}, k = 2, 3, \cdots,$$
$$X_1 = S_1 \quad 和 \quad X_k = S_k - S_{k-1}, k = 2, 3, \cdots.$$

显然, 序列 $X_1, X_2, X_3, \cdots$ 构成一列独立同分布正整数值随机变量. 更新过程 $S_m = X_1 + \cdots + X_m, m \geqslant 1$, 可解释为 U_n 相继超过它前面最大值的时间 (下标). 换句话说, 在这些时间形成了一列新的最大值.

(ii) 序列 $\{U_{S_m} - U_{S_{m-1}}\}_{m=1}^{\infty}$ 是一个 i.i.d. 正随机变量序列, 因此它可以产生一个更新过程. 这个例子在独立随机变量和的波动理论中是重要的 (见第 2 卷第 17 章).

(g) 马尔可夫链中的更新过程

设 $Z_0, Z_1, \cdots$ 是常返的马尔可夫链. 假设 $Z_0 = i$ 并考虑相邻两次到达状态 i 的间隔时间. 具体地, 令

$$X_1 = \min\{n > 0 : Z_n = i\},$$

和

$$X_{k+1} = \min\{n > X_k : Z_n = i\} - X_k, \quad k = 1, 2, \cdots,$$

由于它们都是从同一初始状态 i 开始计算, 马尔可夫性保证了 $X_1, X_2, \cdots$ 是 i.i.d. 随机变量序列. 这样的 $\{X_k\}$ 构成一个更新过程. 这个事实使得我们在第 3 章中可以应用更新理论证明马尔可夫链的基本极限定理.

(h) 自然嵌入的更新过程

自然嵌入的更新过程可以在很多应用概率的领域中出现, 包括分支过程, 保险模型, 群体增长现象, 进化遗传过程, 工程系统, 经济结构等等.

172

5.3 若干特殊更新过程的补充

A. 作为更新过程的泊松过程

如前所述, 参数为 λ 的泊松过程是更新过程, 其事件出现的间隔时间具有指数分布 $F(x) = 1 - \mathrm{e}^{-\lambda x}, x \geqslant 0$. 指数分布的无记忆性 (见 4.2 节) 可用于计算若干与泊松更新过程有关的函数.

(i) *更新函数*

由于 $N(t)$ 具有泊松分布

$$\Pr\{N(t) = k\} = \frac{(\lambda t)^k \mathrm{e}^{-\lambda t}}{k!}, \quad k = 0, 1, \cdots,$$

因此

$$M(t) = E[N(t)] = \lambda t.$$

(ii) 剩余寿命

注意到在时刻 t 剩余寿命超过 x 当且仅当在区间 $(t, t+x]$ 不出现更新 (见图 5-3). 因为泊松过程具有平稳独立增量, 这个事件与在区间 $(0, x]$ 没有出现更新有相同的概率. 用式子表达, 我们有

$$\begin{aligned} \Pr\{\gamma_t > x\} &= \Pr\{N(t+x) - N(t) = 0\} \\ &= \Pr\{N(x) = 0\} = \mathrm{e}^{-\lambda x}. \end{aligned} \tag{3.1}$$

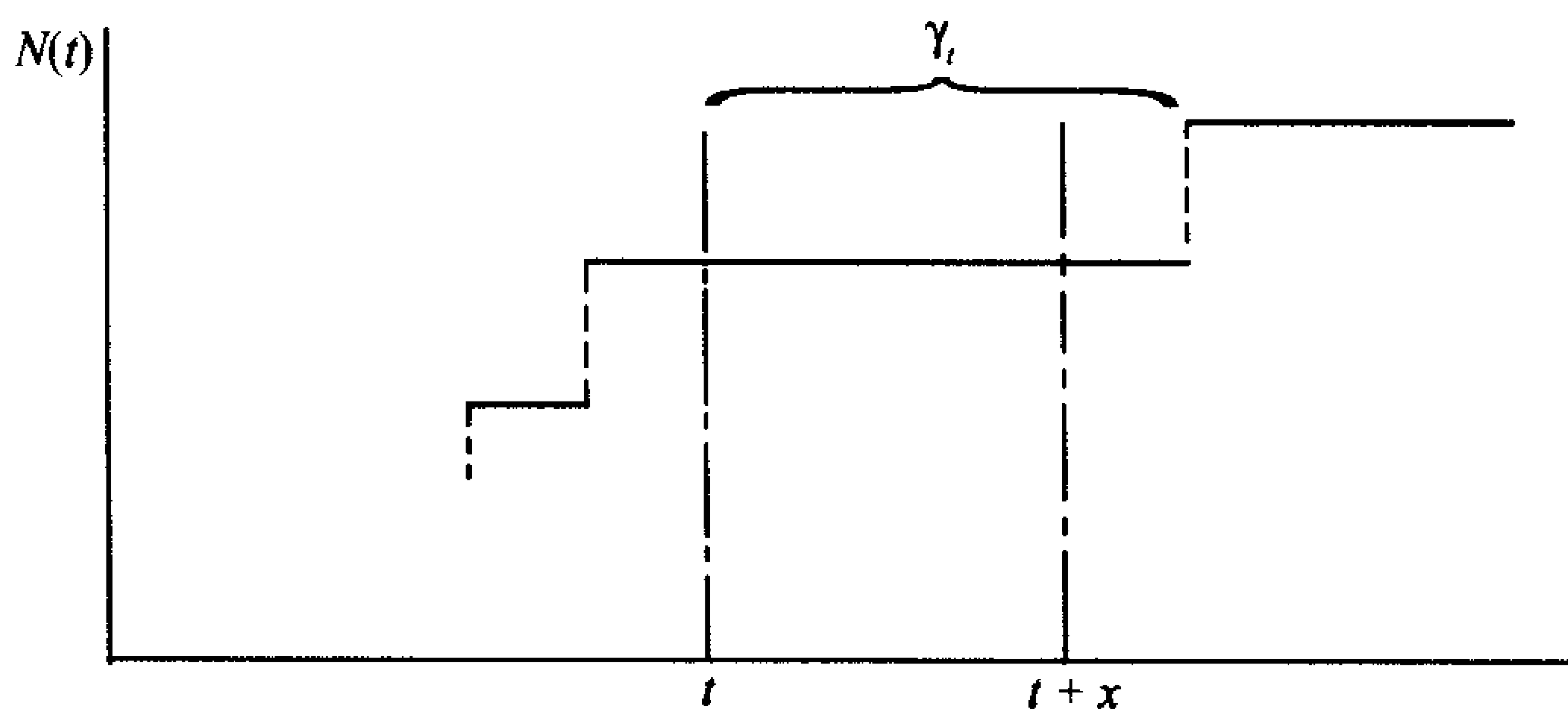


图 5-3

173 这样, 在泊松过程中, 剩余寿命与每个寿命具有相同的指数分布

$$\Pr\{\gamma_t \leqslant x\} = 1 - \mathrm{e}^{-\lambda x}, \quad x \geqslant 0. \tag{3.2}$$

方程 (3.2) 可以写为另一稍为含糊的形式

$$\Pr\{\gamma_t > x\} = 1 - F(x), \quad x \geqslant 0, \quad t > 0. \tag{3.3}$$

在 5.8 节中我们将看到这个恒等式在所有更新过程中刻划了泊松过程.

(iii) 现龄

现龄 δ_t 自然不能超过 t, 而对于 $x < t$, 现龄超过 x 当且仅当区间 $(t-x, t]$ 不出现更新, 这个事件的概率为 $\mathrm{e}^{-\lambda x}$. 因此, 现龄遵循截尾指数分布

$$\Pr\{\delta_t \leqslant x\} = \begin{cases} 1 - \mathrm{e}^{-\lambda x}, & 0 \leqslant x < t, \\ 1, & t \leqslant x. \end{cases} \tag{3.4}$$

为了后面使用方便, 把以上公式写为

$$\Pr\{\delta_t \leqslant x\} = \begin{cases} F(x), & x < t, \\ 1, & x \geqslant t. \end{cases} \tag{3.5}$$

在 5.8 节中将说明, δ_t 的这个性质也刻划了泊松过程.

(iv) 全寿命的均值

利用第 1 章问题 9 关于非负随机变量均值的计算, 我们有

$$\begin{aligned}E[\beta_t] &= E[\gamma_t] + E[\delta_t] \\ &= \frac{1}{\lambda} + \int_0^t \Pr\{\delta_t > x\}\mathrm{d}x \\ &= \frac{1}{\lambda} + \int_0^t \mathrm{e}^{-\lambda x}\mathrm{d}x \\ &= \frac{1}{\lambda} + \frac{1}{\lambda}(1-\mathrm{e}^{-\lambda t}).\end{aligned}$$

174

注意到全寿命均值显著地大于任一更新区间的平均寿命 $1/\lambda = E[X_k]$. 这个现象当 t 较大时, 即过程已经持续运行一段长时间之后, 尤其令人惊讶. 这时全寿命的均值 $E(\beta_t)$ 大约是平均寿命的两倍. 这个事实似乎有点荒谬.

让我们重新检查全寿命 β_t 的定义方式, 并直观解释上述的矛盾. 首先 t 为一个任意固定的时间, 于是 β_t 是测量包含点 t 的更新时间的长度. 这样的方式显然会以较大可能性有利于较长的更新区间而不利于较短的更新区间. 这个现象就是所谓的 "**长度偏倚抽样**". 它虽然很隐蔽, 但在多次抽样下即会呈现.

(v) γ_t 和 δ_t 的联合分布

确定 γ_t 和 δ_t 联合分布的方法和确定 δ_t 分布类似. 事实上, 对任何 $x > 0$ 和 $0 < y < t$, 当且仅当在区间 $(t-y, t+x]$ 不存在更新时, 事件 $\{\gamma_t > x, \delta_t > y\}$ 才出现, 其概率为 $\mathrm{e}^{-\lambda(x+y)}$. 这样,

$$\Pr\{\gamma_t > x, \delta_t > y\} = \begin{cases} \mathrm{e}^{-\lambda(x+y)}, & 若\ x > 0, 0 < y < t, \\ 0, & 若\ y \geqslant t. \end{cases} \tag{3.6}$$

对于泊松过程, γ_t 和 δ_t 是独立的, 因为它们联合分布是边缘分布的乘积. 在更新过程中这个性质也刻划了泊松过程. 这个命题已收在本章末问题 25 中.

B. 替换模型

设 $X_1, X_2, \cdots$ 表示个体 (灯泡、集成电路块和机器等) 的寿命, 这些个体在使用中相继被替换, 当前面一个个体坏了立即换上另一个个体开始使用. 我们约定 $\{X_i\}$ 是具有有限均值 $\mu = E[X_k]$ 的独立同分布正随机变量序列. 由于每个个体平均持续 μ 时间单位, 我们希望在长时间运转中, 每单位时间平均替换 $1/\mu$ 个个体. 即, 我们希望

$$\frac{1}{t}M(t) \to \frac{1}{\mu}, \quad 当\ t \to \infty. \tag{3.7}$$

这个事实将在 5.4 节论证.

175

在长期运转中, 任何在个体损坏之前就被替换的策略在每单位时间使用的个体将多于 $1/\mu$. 然而, 在使用中避免损坏在经济上有好处, 元部件往往是随着使用时间 (依龄) 而退化的, 因此考虑选择替换策略可能更有利.

一个个体或损坏或到达某使用时间 T, 无论哪一个先出现就被替换称为**依龄替换策略**. 直观地说, 在长时间的运行中, 由于损坏 (发生在 T 之前) 被替换的个体所占比例为 $F(T)$, 而对应于在 T 时刻被替换的个体 (想象并不太昂贵) 所占比例为 $1-F(T)$. 对于这个依龄替换策略模型, 更新间距显然服从分布律

$$F_T(x)=\begin{cases} F(x), & x<T, \\ 1, & x\geqslant T, \end{cases}$$

其平均更新间距为

$$\mu_T=\int_0^\infty\{1-F_T(x)\}\mathrm{d}x=\int_0^T\{1-F(x)\}\mathrm{d}x<\mu.$$

用导出关系式 (3.7) 相同的理由可证明长期运行中平均替换率递增趋向于 $1/\mu_T$.

现在令 $Y_1, Y_2, \cdots$ 表示个体真正相继损坏的间隔时间. 随机变量 Y_1 是由随机数个长度为 T 的时间区间 (对应于与损坏无关的替换) 加上最后的时间区间段组成, 最后时间区间段的分布是在 T 前损坏的条件下的损坏分布, 即 Y_1 具有 $NT+Z$ 的分布, 其中

$$\Pr\{N\geqslant k\}=\{1-F(T)\}^k, \quad k=0,1,\cdots,$$

且

$$\Pr\{Z\leqslant z\}=F(z)/F(T), \quad 0\leqslant z\leqslant T.$$

因此,

$$\begin{aligned} E[Y_1]&=\frac{1}{F(T)}\left\{T[1-F(T)]+\int_0^T(F(T)-F(x))\mathrm{d}x\right\} \\ &=\frac{1}{F(T)}\int_0^T\{1-F(x)\}\mathrm{d}x. \end{aligned}$$

个体真正损坏的间隔时间随机变量序列 $\{Y_i\}$ 产生一个更新过程, 在长期运行中, 每单位时间平均损坏率是 $1/E[Y_1]$. 这个结论仍然是依据导出关系式 (3.7) 的相同理由. 由于 $E[Y_1]$ 依赖于 F, 依龄替换策略的损坏率 $1/E(Y_1)$ 可能小于仅当损坏时才替换的损坏率 $1/\mu$.

176

在下面情况可考虑使用整体替换策略：若干个元件任何时刻都同时使用, 同时更换所有的元件要比个别地替换元件在经济上更节约. 整体替换要求, 或者替换个别损坏的元件或者在时刻 $T, 2T, 3T, \cdots$ 替换所有元件. 从而, 在每个 T 单位时间进行一次有计划的或整体的替换和 (平均)M(T) 次损坏替换. 这样, 长期运行中每单位时间更换次数是 $(1+M(T))/T$. 可把它与其他替换策略相比较.

C. 计数器模型

计数器是探测和记录瞬时脉冲信号的仪器. Geiger-Muller 计数器就是大家熟悉的例子, 它是探测大气层宇宙幅射或者有源幅射, 例如由一块镭放射出来的射线. 另一个例子是电子扩程器.

所有物理上可实现计数器都是不完善的, 无能力检测进入仪器的所有信号. 当一个粒子或信号被记录下来以后, 计数器必须恢复到原来状态以准备记录下一个信号. 在仪器调整时间 (称为闭锁时间或断开时间) 到达的信号无法记录. 我们必须区分到达的粒子和被记录的粒子. 仪器仅能探测被记录的粒子. 人们希望由此推断到达过程的性质.

假设粒子或者信号到达的过程是更新过程, 其间隔时间为 $X_1, X_2, \cdots$, 且 $\Pr\{X_k \leqslant x\} = F(x)$. 计数器模型依机构闭锁时间的种类的差别而不同. 下面我们讨论其中最普通的两种类型.

I 型计数器

一个粒子在时刻 0 到达, 并且使计数器闭锁一段时间 Y_1. 第一个被记录的粒子是在时刻 Y_1 之后到达的第一个粒子. 由于记录粒子, 计数器闭锁, 其时间长度为 Y_2. 下一次被记录的粒子是在计数器开放以后到达的第一个粒子. 这个过程不断重复, 其中相继闭锁时间记为 $Y_1, Y_2, Y_3, \cdots$, 假定它们是独立的且具有相同的分布 $\Pr\{Y_k \leqslant y\} = G(y)$, 并且与到达过程 $\{X_k\}$ 互相独立.

177

令 Z_1 表示直到第一个信号 (不计算在时刻 0 到达的一个) 被记录的时间, 令 $Z_n, n = 2, 3 \cdots$ 表示第 $n-1$ 次记录和第 n 次记录之间的时间间距. 由于过程在每次记录后又重新开始, 所以 $\{Z_k\}$ 构成一个更新过程. 以上所述的情况如图 5-4 所示.

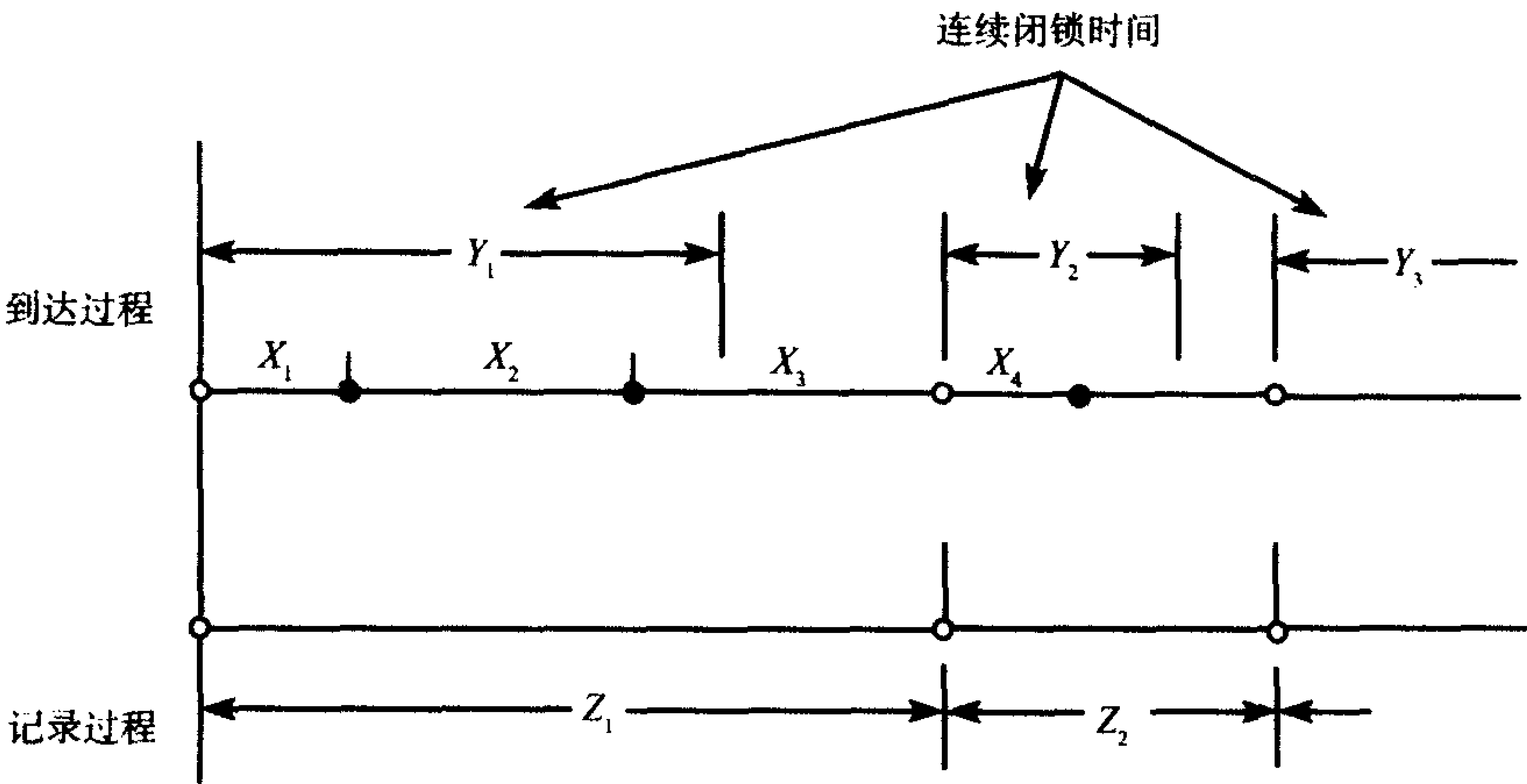


图 5-4 I 型计数器, ● 表示未被记录的信号, ○ 表示被记录的信号

受图 5-4 的启发, 知道 Z_1 是 Y_1 加上在 Y_1 的剩余寿命, 或

$$Z_1 = Y_1 + \gamma_{Y_1} = S_{N(Y_1)+1},$$

其中 $S_k = X_1 + X_2 + \cdots + X_k$.

由于过程 $\{X_n\}$ 和 $\{Y_n\}$ 是独立的, 借助于全概率公式我们得到

$$\begin{aligned}\Pr\{Z_1 \leqslant z\} &= \int_0^z \Pr\{y + \gamma_y \leqslant z | Y_1 = y\} \mathrm{d}G(y) \\ &= \int_0^z \{1 - A_{z-y}(y)\} \mathrm{d}G(y),\end{aligned}$$

其中 $A_x(t) = \Pr\{\gamma_t > x\}$.

$A_x(t)$ 的显示公式可从后面方程 (6.1) 得到, 所以 I 型计数器计数间隔时间的分布是完全确定的.

在长时间运行中, 每单位时间内平均记录 $1/E(Z_1)$ 个粒子, 而平均到达 $1/E[X_1]$ 个粒子. 后面我们会说明, 在长期运行中, 在所有到达粒子中被记录粒子的比例是 $E[X_1]/E[Z_1]$.

178

当到达过程是参数为 λ 的泊松过程时, 事件出现的间隔时间的指数分布无记忆性告诉我们, γ_t 遵循指数分布且与 t 无关. 这样, 在泊松情况下,

$$\Pr\{Z_1 \leqslant z\} = \int_0^z G(z - y)\lambda \mathrm{e}^{-\lambda y} \mathrm{d}y. \tag{3.8}$$

II 型计数器

这里的闭锁机制更加复杂. 如前, 一个输入信号被记录当且仅当该信号是在计数器开放时间到达. 然而在前面例子中只有被记录的粒子能使计数器闭锁. 对于 II 型计数器, 每个到达的信号都能延长计数器闭锁时间, 相连结的闭锁时间同时叠加起来. 例如, 假设第一个粒子闭锁计数器的时间长度为 σ_1, 而第二个脉冲在时间 $\tau < \sigma_1$ 到达. 并独立造成长度为 σ_2 的闭锁时间, 那么计数器下次开放时间是在时刻 σ_1 或 $\tau + \sigma_2$, 无论哪种情况出现, 我们都应假定在此时刻之前没有另外粒子到达 [1]. 这个过程的一个典型实现如图 5-5 所示.

如 I 型计数器, 设 Z_n 是第 $n-1$ 次被记录脉冲和第 n 次被记录脉冲之间的时间长度, $\{Z_k\}$ 也是一个更新过程.

对这种计数器过程进行一般分析是十分困难的, 在到达过程是参数为 λ 的泊松过程的假设之下, 我们得出一些结果.

1. 应假设在此时刻之前没有别的粒子到达或到达粒子所造成的闭锁时间在时刻 σ_1 或 $\tau + \sigma_2$ 之前结束. —— 译者注

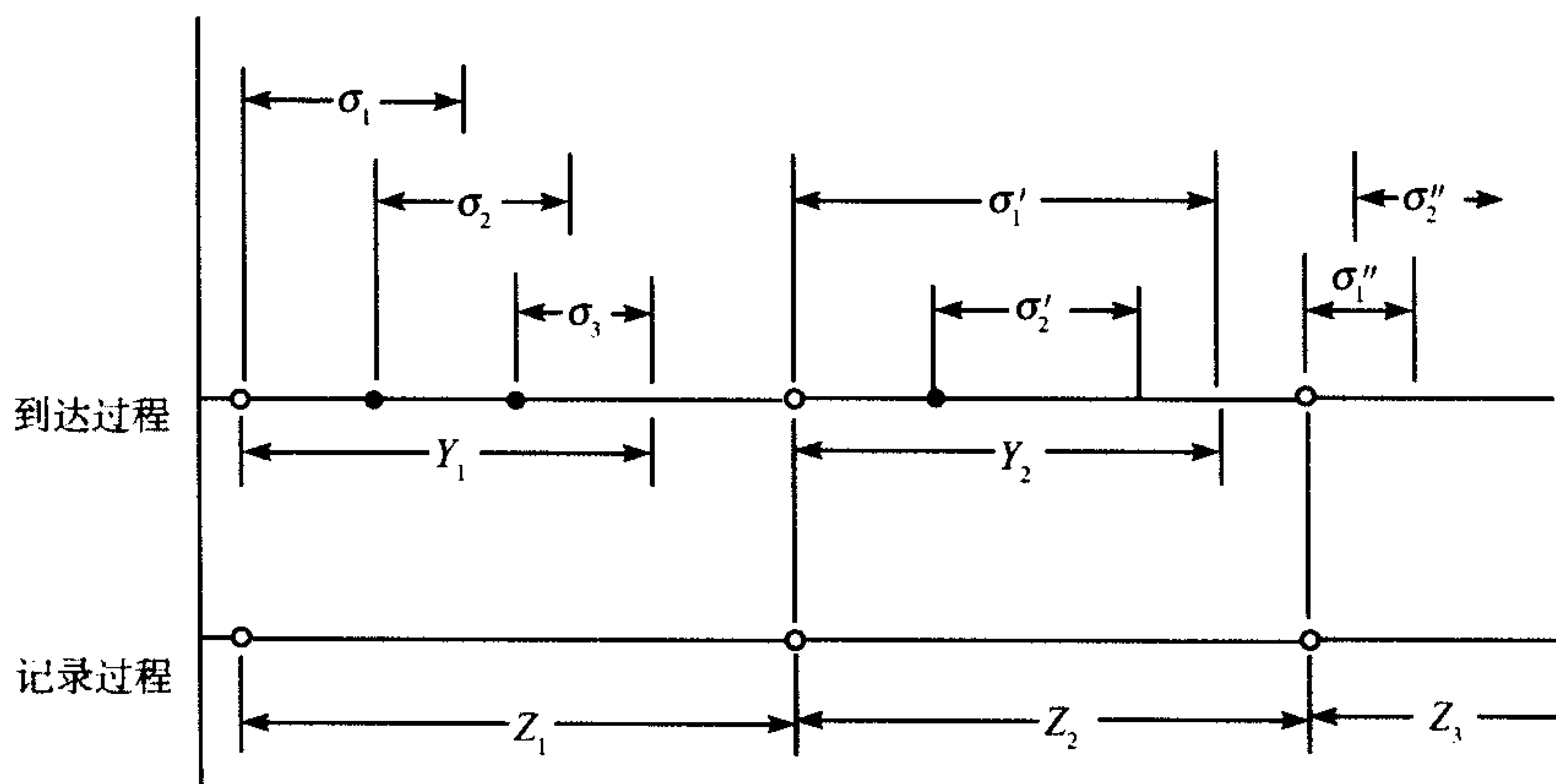


图 5-5 II 型计数器, • 表示未被记录的信号, ∘ 表示被记录的信号

179

设 $p(t)$ 是计数器在时刻 t 开放的概率. 我们下面阐明

$$p(t)=\exp\left\{-\lambda\int_0^t[1-G(y)]\mathrm{d}y\right\},\tag{3.9}$$

其中 $G(y)=\Pr\{\sigma_k\leqslant y\}$. 为了导出这个公式, 回忆第 4 章定理 2.3, 已知泊松过程在时间区间 $(0,t]$ 出现 n 次, 其出现时间的分布相当于从 $(0,t]$ 上均匀分布的总体中抽取出的 n 个独立随机变量的分布. 计数器在时刻 t 是开放的, 当且仅当由这 n 个信号产生的所有闭锁时间在时刻 t 之前结束. 一个闭锁时间区间从时刻 y 开始, 在时刻 t 前结束的概率是 $G(t-y)$. 倘若一个信号在时间区间 $(0,t]$ 内出现, 它的实际到达时间具有均匀分布. 因此由它导致的闭锁时间在时刻 t 之前结束具有概率 $\int_0^t G(t-y)\mathrm{d}y/t$. 由于闭锁时间假定是独立的并且与到达过程无关, 我们有

$$\begin{aligned}&\Pr\{\text{计数器在时刻 } t \text{ 开放}|n \text{ 个信号出现在 } (0,t]\}\\&=\left\{\int_0^t G(t-y)\frac{1}{t}\mathrm{d}y\right\}^n.\end{aligned}$$

但信号在时间区间 $(0,t]$ 到达的数目具有均值为 λt 的泊松分布. 借助于全概率公式, 我们得到

$$\begin{aligned}p(t)&=\sum_{j=0}^{\infty}\left\{\int_0^t G(t-y)\frac{1}{t}\mathrm{d}y\right\}^j\frac{(\lambda t)^j\mathrm{e}^{-\lambda t}}{j!}\\&=\exp\left\{-\lambda t\left[1-\int_0^t G(t-y)\frac{1}{t}\mathrm{d}y\right]\right\}\\&=\exp\left\{-\lambda\int_0^t[1-G(y)]\mathrm{d}y\right\},\end{aligned}$$

公式 (3.9) 得证.

继续假定信号流是泊松的, 记 $M_R(t)$ 为在 $(0,t]$ 上被记录信号的平均数, 则 $p(t)$ 与 $M_R(t)$ 有关. 我们证明

$$\frac{\mathrm{d}M_R(t)}{\mathrm{d}t} = \lambda p(t). \tag{3.10}$$

180 为证明上式, 首先阐述下面相关论断.

信号出现在区间 $(t,t+h]$ 的概率是 $\lambda h + o(h)$. 由定义, 在相同的短时间区间里发现计数器开放的概率为 $p(t)+o(h)$. 所以除了 $o(h)$ 项外, $\lambda hp(t)$ 是某个信号在 $(t,t+h]$ 被记录的概率. 既然在时间区间 $(t,t+h]$ 出现至少两个信号的概率是 $o(h)$, 我们有

$$M_R(t+h) = M_R(t) + \lambda hp(t) + o(h),$$

其中 $o(h)$ 合并了所有可忽略的项. 则

$$\frac{\mathrm{d}M_R(t)}{\mathrm{d}t} = \lim_{h\downarrow 0}\frac{M_R(t+h)-M_R(t)}{h} = \lambda p(t),$$

此即 (3.10).

注意, 由其原意, $M_R(0)=0$. 现在结合 (3.9) 和 (3.10), 然后求积分得

$$M_R(t) = \int_0^t \lambda \exp\left\{-\lambda\int_0^s [1-G(y)]\mathrm{d}y\right\}\mathrm{d}s. \tag{3.11}$$

5.4 更新方程和初等更新定理

A. 更新函数

在 5.1 节中我们导出在 $(0,t]$ 上平均计数次数的公式:

$$M(t) = E[N(t)] = \sum_{j=1}^{\infty} F_j(t), \tag{4.1}$$

其中

$$F_j(t) = \Pr\{S_j \leqslant t\}, \quad t \geqslant 0.$$

由于它的深刻含义及关系已超过它作为平均计数次数的解释, 人们给予 $M(t)$ 一个专门的名称 —— 更新函数 [1].

1. 一些作者曾试图定义含原点的更新过程, 并定义更新函数为 $1+M(t)=E[1+N(t)]=\sum_{n=0}^{\infty}F_n(t)$, 其中 $F_0(t)=1$ 对于 $t\geqslant 0$, 其他为 0. 这个定义可简化某些公式, 但在勒贝格 – 斯蒂尔切斯积分理论方面要麻烦些.

我们的第一件事是说明, 对于每个 $t>0$, $M(t)$ 是有限的. 为此, 首先根据定义和 $F_j(x)$ 的单调性, 推出不等式

$$F_n(t)=\int_0^t F_{n-m}(t-\xi)\mathrm{d}F_m(\xi)\leqslant F_{n-m}(t)F_m(t),\quad 1\leqslant m\leqslant n-1.$$

181

特别, 对任何整数 k,r 和 n, 我们有

$$F_{nr+k}(t)\leqslant F_{(n-1)r+k}(t)F_r(t).$$

直接迭代导出关系式

$$F_{nr+k}(t)\leqslant [F_r(t)]^n F_k(t),\quad 0\leqslant k\leqslant r-1. \tag{4.2}$$

由于 (4.2), 可见若存在 r, 使得 $F_r(t)<1$, 则级数 $M(t)=\sum_{k=1}^{\infty}F_k(t)$ 收敛, 并至少依几何级数速度收敛.

因为 X_i 是正随机变量, 所以 $F(0+)=0$, 从而存在某个正数 t_0 使得 $F(t_0)<1$, 由此可归纳地推断对于每个 $t>0$, 必定存在 r 满足 $F_r(t)<1$. 这个事实从概率意义考虑是直观的, 并且等价于如下命题: 独立同分布正随机变量序列的部分和 S_n 以概率 1 递增趋向于无限.

由前面的讨论引出两个重要的推论, 为便于引用, 现列在下面:

对任意固定 t

$$F_n(t)\to 0,\ \text{当}\ n\to\infty\ (\text{至少以几何级数速度收敛}), \tag{4.3}$$

和

$$M(t)<\infty,\quad \text{对所有}\ t.$$

显然, 由其意义 (或由公式 (4.1))$M(t)$ 是关于 t 的不减函数. 此外, 易验证 $M(t)$ 是右连续的 (由于每个 $F_k(t)$ 具有这个性质并且级数在有限区间一致收敛). 这样 $M(t)$ 有分布函数的特征, 只不过 $M(\infty)=\lim_{t\to\infty}M(t)\neq 1$ (实际上, $M(\infty)=\infty$). 从而可以与 $\int a(t-y)\mathrm{d}F(y)$ 平行地定义表达式 $\int a(t-y)\mathrm{d}M(y)$. 特别, 当 M 可微分时, 若 $m(t)=\mathrm{d}M(t)/\mathrm{d}t$, 则积分 $\int a(t-y)\mathrm{d}M(y)$ 简化为 $\int a(t-y)m(y)\mathrm{d}y$.

在此观点之下, 我们可推广卷积的概念至任何两个增函数. 设 A 和 B 是不减、右连续函数, 且 $A(0)=B(0)=0$. A 和 B 的卷积, 记为 $A*B$, 定义如下,

$$A*B(t)=\int_0^t B(t-y)\mathrm{d}A(y),\quad t\geqslant 0. \tag{4.4}$$

182

由于 $B(0)=0$, 我们有 $B(t-y)=\int_0^{t-y}\mathrm{d}B(z)$. 把此式代入 (4.4), 改变积分的顺序, 得

$$\begin{aligned}A*B(t)&=\int_0^t\left\{\int_0^{t-y}\mathrm{d}B(z)\right\}\mathrm{d}A(y)\\&=\int_0^t\left\{\int_0^{t-z}\mathrm{d}A(y)\right\}\mathrm{d}B(z)\\&=B*A(t),\end{aligned}$$

因此 “$*$” 是一个可交换运算.

现在我们证明更新函数 $M(t)$ 满足方程

$$M(t)=F(t)+\int_0^t M(t-y)\mathrm{d}F(y),\quad t\geqslant 0,$$

或以卷积记号记为

$$M(t)=F(t)+F*M(t),\quad t\geqslant 0. \tag{4.5}$$

这个恒等式借助于**更新论证法**得以证实. 该方法是通过假定第一次更新的时间为 X_1, 然后计算更新的期望数. 显然, 过程的概率结构在时刻 X_1 之后又重新开始, 因此

$$E[N(t)|X_1=x]=\begin{cases}0, & \text{若 } x>t,\\ 1+M(t-x), & \text{若 } x\leqslant t.\end{cases}$$

换句话说, 如果第一次寿命 X_1 超过 t, 则在区间 $(0,t]$ 没有更新. 另一方面, 当 $X_1=x<t$, 在 x 处存在一次更新并在 $(x,t]$ 按平均意义有 $M(t-x)$ 次更新. 应用全概率公式有

$$\begin{aligned}M(t)&=E[N(t)]\\&=\int_0^t E[N(t)|X_1=x]\mathrm{d}F(x)\\&=\int_0^t\{1+M(t-x)\}\mathrm{d}F(x)\\&=F(t)+\int_0^t M(t-x)\mathrm{d}F(x),\end{aligned}$$

183 于是关系式 (4.5) 得以证实.

更新理论的许多活力均源于上述的推理方法, 即研究从第一个 “事件” 出现的时刻重新开始的动态更新过程.

B. 更新方程

以下形式的积分方程

$$A(t)=a(t)+\int_0^t A(t-x)\mathrm{d}F(x),\quad t\geqslant 0,\tag{4.6}$$

称为更新方程. 其中 $a(t)$ 和分布函数 $F(t)$ 是已知的, 而未知量为 $A(t)$.

若无不明确, 我们将使用记号 $B*C(t)$ 表示函数 $C(t)$(假定适当光滑且在有限区间是有界的) 与右连续增函数 $B(t)$ 的卷积, 其中 $B(0)=0$, 用式子表达如下:

$$B*C(t)=\int_0^t C(t-\tau)\mathrm{d}B(\tau).\tag{4.7}$$

若 $B'(t)=b(t)$ 存在, 则 (4.7) 可化为

$$B*C(t)=\int_0^t C(t-\tau)b(\tau)\mathrm{d}\tau,$$

这里, 假设后面的积分在通常意义下存在. 关于 $B*C$ 一些初等性质我们列出下面几条以供参考, 它们的证明比较容易, 留给读者自己完成.

(i) $\max\limits_{0\leqslant t\leqslant T}|(B*C)(t)|\leqslant\max\limits_{0\leqslant t\leqslant T}|C(t)|\cdot B(T)$ (也可参考下面 (4.11));

(ii) $B*C_1+B*C_2=B*(C_1+C_2)$;

(iii) 如果 B_1 和 B_2 是递增的, 则

$$B_1*(B_2*C)=(B_1*B_2)*C.\tag{4.8}$$

比较 (4.6) 与 (4.5), 得知更新函数 $M(t)$ 满足当 $a(t)=F(t)$ 时的更新方程. 下面定理断言一个任意更新方程的解可以用更新函数来表示.

定理 4.1 假设 a 是有界函数, 则存在唯一的一个函数 A, 它在有限区间是有界的, 并且满足

$$A(t)=a(t)+\int_0^t A(t-y)\mathrm{d}F(y).\tag{4.9}$$

184

这个函数是

$$A(t)=a(t)+\int_0^t a(t-x)\mathrm{d}M(x),\tag{4.10}$$

其中 $M(t)=\sum\limits_{k=1}^{\infty}F_k(t)$ 是更新函数.

证明 我们首先证明由 (4.10) 确定的函数 A 满足所要求的有界性并且实际上就是 (4.9) 的解. 因为 a 是有界函数而 M 是不减和有限的, 对每个固定 T 推得

$$\begin{aligned}\sup_{0\leqslant t\leqslant T}|A(t)| &\leqslant \sup_{0\leqslant t\leqslant T}|a(t)| + \int_0^T \Big\{\sup_{0\leqslant y\leqslant T}|a(y)|\Big\}\mathrm{d}M(x) \\ &= \sup_{0\leqslant t\leqslant T}|a(t)|\{1+M(T)\}<\infty,\end{aligned} \tag{4.11}$$

这就证明了表达式 (4.10) 在有限区间是有界的. 为检验 (4.10) 的 $A(t)$ 满足 (4.9), 我们有

$$\begin{aligned}A(t) &= a(t) + M*a(t) \\ &= a(t) + \Big(\sum_{k=1}^{\infty}F_k\Big)*a(t) \\ &= a(t) + F*a(t) + \sum_{k=2}^{\infty}F_k*a(t) \\ &= a(t) + F*\Big\{a(t)+\Big(\sum_{k=1}^{\infty}F_k\Big)*a(t)\Big\} \quad (\text{因为 } F_k = F*F_{k-1} \text{ 并利用 (4.8)}) \\ &= a(t) + F*A(t).\end{aligned}$$

为完成定理 4.1 的证明, 接下来是验证 A 的唯一性. 这只须证明更新方程 (4.9) 的任何解在有限区间是有界的, 可用 (4.10) 表示即可.

为此, 注意到更新方程 (4.9) 适于使用逐次逼近法, 将 $A(t)$ 的表达式重复代入 (4.9) 的右边, 并适当地展开. 我们利用卷积的记号实现这个步骤. 把 (4.9) 缩写成

$$A = a + F*A$$

把 A 代入到上式右边, 得

$$\begin{aligned}A &= a + F*(a+F*A) \\ &= a + F*a + F*(F*A) \\ &= a + F*a + F_2*A \quad (\text{由于恒等式 } F_2 = F*F),\end{aligned}$$

185

这里, 心照不宣地利用了性质 (4.8). 我们重复这个步骤, 得到等式

$$\begin{aligned}A &= a + F*a + F_2*(a+F*A) \\ &= a + F*a + F_2*a + F_3*A = \cdots \\ &= a + \Big(\sum_{k=1}^{n-1}F_k\Big)*a + F_n*A.\end{aligned}$$

其次, 注意到

$$|F_n * A(t)| = \left|\int_0^t A(t-y)\mathrm{d}F_n(y)\right| \leqslant \left\{\sup_{0\leqslant y\leqslant t}|A(t-y)|\right\} \times F_n(t).$$

既然 A 在有限区间是有界的, 并且 $\lim_{n\to\infty} F_n(t) = 0$(参考式 (4.3)), 由此推得对每个固定 t 有 $\lim_{n\to\infty}|F_n * A(t)| = 0$. 类似地, 由于 a 有界, 得

$$\lim_{n\to\infty}\left(\sum_{k=1}^{n-1} F_k\right) * a(t) = \left(\sum_{k=1}^{\infty} F_k\right) * a(t) = M * a(t).$$

同样,

$$A(t) = a(t) + \lim_{n\to\infty}\left\{\sum_{k=1}^{n-1} F_k * a(t) + F_n * A(t)\right\} = a(t) + M * a(t),$$

故方程 (4.9) 的一般解 A 具有表达式 (4.10). 这就完成了定理 4.1 的唯一性证明. ■

借助于定理 4.1, 我们来证明重要关系式

$$\begin{aligned} E[S_{N(t)+1}] &= E[X_1 + X_2 + \cdots + X_{N(t)+1}] \\ &= E[X_1]\cdot E[N(t)+1] \\ &= E[X_1]\cdot[M(t)+1], \end{aligned} \tag{4.12}$$

初看起来, 这个恒等式有点像第 1 章所导出的一个恒等式, 后者表明若 $X_1, X_2, X_3, \cdots$ 是独立同分布随机变量序列. N 是与 X_i 无关的整数值随机变量, 且它们的均值都存在, 则等式 $E[X_1 + \cdots + X_n] = E[X_1]\cdot E[N]$ 成立. 等式 (4.12) 与此式子关键差别在于求和数目 $N(t)+1$ 不独立于被加项. 例如, 回忆在 5.3 节讨论的泊松过程, 全寿命 β_t 是含于 $S_{N(t)+1}$ 的最后被加项, 当 t 足够大时其均值大约是无条件均值的两倍. 由于这个理由, 把 $E[S_{N(t)}]$ 作为 $E[X_1]$ 和 $E[N(t)]$ 的乘积来求值是不正确的. 鉴于这些引以为诫的说明, 等式 (4.12) 更加有意思. (它实际是著名的 Wald 恒等式的一个特别情况, 可参考第 6 章有关部分.) 186

为了导出 (4.12), 我们利用更新论证法构造关于 $A(t) = E[S_{N(t)+1}]$ 的更新方程. 照例, 我们假设第一次更新时间是 $X_1 = x$, 然后区别两种可能性: 第一种情况是 $x > t$, 所以 $N(t) = 0$ 并且 $S_{N(t)+1} = x$, 第二种情况是 $x \leqslant t$. 通过对所包含的量直接解释易于验证等式:

$$E[S_{N(t)+1}|X_1 = x] = \begin{cases} x, & \text{若 } x > t, \\ x + A(t-x), & \text{若 } x \leqslant t. \end{cases}$$

其次, 援引全概率公式得

$$\begin{aligned}A(t) &= E[S_{N(t)+1}]\\&= \int_0^\infty E[S_{N(t)+1}|X_1 = x]\mathrm{d}F(x)\\&= \int_0^t [x + A(t-x)]\mathrm{d}F(x) + \int_t^\infty x\mathrm{d}F(x)\\&= \int_0^\infty x\mathrm{d}F(x) + \int_0^t A(t-x)\mathrm{d}F(x)\\&= E[X_1] + \int_0^t A(t-x)\mathrm{d}F(x).\end{aligned}$$

这样 $A(t) = E[S_{N(t)+1}]$ 满足更新方程, 其中 $a(t) =$ 常数 $E[X_1]$. 定理 4.1 说明

$$\begin{aligned}A(t) &= a(t) + \int_0^t a(t-x)\mathrm{d}M(x)\\&= E[X_1] + \int_0^t E[X_1]\mathrm{d}M(x)\\&= E[X_1] \times [1 + M(t)],\end{aligned}$$

187

由此证明了 (4.12)

注意到剩余寿命 $\gamma_t = S_{N(t)+1} - t$ 有

$$E[\gamma_t] = E[X_1] \cdot [1 + M(t)] - t. \tag{4.13}$$

在 5.3 节中多次使用了直观结果 $M(t)/t \to 1/\mu$, $t \to \infty$, 其中 $\mu = E[X_1]$. 现在我们已能充分证明这个重要论断, 它通常被看作是**初等更新定理**.

定理 4.2 设 $\{X_i\}$ 是更新过程, $\mu = E[X_1] < \infty$, 则

$$\lim_{t\to\infty} \frac{1}{t}M(t) = \frac{1}{\mu}.$$

证明 由于情况 $t < S_{N(t)+1}$ 总是成立的, 结合等式 (4.12), 我们有

$$t < E[S_{N(t)+1}] = \mu[1 + M(t)],$$

因此

$$\frac{1}{t}M(t) > \frac{1}{\mu} - \frac{1}{t}.$$

由此推得

$$\liminf_{t\to\infty} \frac{1}{t}M(t) \geqslant \frac{1}{\mu}. \tag{4.14}$$

为了构造相反的不等式, 令 $c>0$ 是任意选定的, 置

$$X_i^c=\begin{cases} X_i, & 若\ X_i\leqslant c, \\ c, & 若\ X_i>c, \end{cases}$$

并考虑具有寿命 $\{X_i^c\}$ 的更新过程.

令 S_n^c 和 $N^c(t)$ 分别表示由 $\{X_i^c\}$ 产生的截尾更新过程的等候时间和计数过程. 由于随机变量 X_i^c 一致有界于 c, 易知有 $t+c\geqslant S_{N^c(t)+1}^c$, 所以

$$t+c\geqslant E[S_{N^c(t)+1}^c]=\mu^c[1+M^c(t)],$$

其中

$$\mu^c=E[X_i^c]=\int_0^c\{1-F(x)\}\mathrm{d}x,$$

且

$$M^c(t)=E[N^c(t)].$$

188

显然, 由 $X_i^c\leqslant X_i$, 可知 $N^c(t)\geqslant N(t)$, 因而 $M^c(t)\geqslant M(t)$. 由此推得

$$t+c\geqslant\mu^c[1+M(t)],$$

重新整理上式得

$$\frac{1}{t}M(t)\leqslant\frac{1}{\mu^c}+\frac{1}{t}\left(\frac{c}{\mu^c}-1\right).$$

因此对任意 $c>0$ 有

$$\limsup_{t\to\infty}\frac{1}{t}M(t)\leqslant\frac{1}{\mu^c}. \tag{4.15}$$

由于

$$\begin{aligned}\lim_{c\to\infty}\mu^c&=\lim_{c\to\infty}\int_0^c[1-F(x)]\mathrm{d}x\\&=\int_0^\infty[1-F(x)]\mathrm{d}x=\mu,\end{aligned}$$

而式 (4.15) 的左边部分是固定的, 由此我们推出

$$\limsup_{t\to\infty}\frac{1}{t}M(t)\leqslant\lim_{c\to\infty}\frac{1}{\mu^c}=\frac{1}{\mu}. \tag{4.16}$$

结合不等式 (4.14) 和 (4.16) 得

$$\lim_{t\to\infty}\frac{1}{t}M(t)=\frac{1}{\mu},$$

从而定理证毕. ■

5.5 更新定理

本节的主题是介绍应用概率论中的一个非常基本的定理即更新定理. 此定理改进了由定理 4.2 所构造的渐近关系式 $M(t) \sim t/\mu, t \to \infty$.

更新定理可理解为极限公式 $\lim\limits_{t\to\infty} M(t)/t = 1/\mu$ 的微分形式. 明确而言, 倘若对 $F(x)$ 附加一些适当条件, 则由更新定理可断言, 对任意 $h > 0$ 有

189

$$M(t+h) - M(t) \to \frac{h}{\mu}, \quad \text{当 } t \to \infty. \tag{5.1}$$

换言之, 当过程经过很长一段时间, 在长度为 h 的区间上更新次数的期望值近似于 h/μ. 虽然定理 5.1 的结论有多种形式, 但都等价于 (5.1), 它不仅提供了应用更新定理的方便形式. 此外, 更新定理在确定更新方程解的渐近特征时有着重要作用.

更新定理的证明比较冗长但又是必要的. 我们略去其细节部分, 并推荐读者参看 Feller[1] 以了解证明细节过程. 然而, 我们将相当用心地叙述以使得读者能理解它的意义, 并能迅速和准确地应用它. 在随后各节里有大量应用, 基本更新定理的含义将变得更加清楚易懂. 为了叙述准确, 我们需要几个预备定义. (只对应用感兴趣的读者可粗略阅读本节后面内容).

定义 5.1 点 α 称为分布函数 F 的**增点**, 如果对任意正数 ε 有

$$F(\alpha + \varepsilon) - F(\alpha - \varepsilon) > 0.$$

分布函数称为是**算术的**, 如果存在一个正数 λ 使得 F 的增点仅仅出现在点 $0, \pm\lambda, \pm 2\lambda, \cdots$ 之中, 具有此性质的最大 λ 称为 F 的**跨距**.

有连续部分的分布函数 F 不是算术的. 具有可能值 $0, 1, 2, \cdots$ 的离散随机变量的分布函数是算术的并且跨距为 1.

定义 5.2 设 g 是定义在 $[0, \infty)$ 上的函数. 对每个正数 δ 和 $n = 1, 2, \cdots$, 令

$$\underline{m}_n = \min\{g(t) : (n-1)\delta \leqslant t \leqslant n\delta\},$$

$$\overline{m}_n = \max\{g(t) : (n-1)\delta \leqslant t \leqslant n\delta\},$$

$$\underline{\sigma}(\delta) = \delta \sum_{n=1}^{\infty} \underline{m}_n, \quad \text{和 } \overline{\sigma}(\delta) = \delta \sum_{n=1}^{\infty} \overline{m}_n.$$

如果级数 $\underline{\sigma}(\delta)$ 和 $\overline{\sigma}(\delta)$ 对每个 δ 是绝对收敛的, 并且当 $\delta \to 0$ 时, 差 $\overline{\sigma}(\delta) - \underline{\sigma}(\delta)$ 趋向于 0, 则称 g 为**直接黎曼可积的**.

每个单调函数 g 若在如下意义下绝对可积

190

$$\int_0^{\infty} |g(t)| \mathrm{d}t < \infty, \tag{5.2}$$

则必定是**直接黎曼可积的**. 这对于我们而言是最重要的情况. 显然, 所有满足 (5.2) 的单调函数的有限线性组合也是直接黎曼可积的.

定理 5.1(基本更新定理) 设 F 是均值为 μ 的正随机变量的分布函数. 设 a 是直接黎曼可积的, A 是更新方程

$$A(t) = a(t) + \int_0^t A(t-x)\mathrm{d}F(x). \tag{5.3}$$

的解.

(i) 如果 F 不是算术的, 则

$$\lim_{t\to\infty} A(t) = \begin{cases} \dfrac{1}{\mu}\displaystyle\int_0^\infty a(x)\mathrm{d}x, & 若\ \mu < \infty, \\ 0, & 若\ \mu = \infty. \end{cases}$$

(ii) 若 F 是算术的且跨距为 λ, 则对所有 $0 \leqslant c < \lambda$ 有

$$\lim_{n\to\infty} A(c+n\lambda) = \begin{cases} \dfrac{\lambda}{\mu}\displaystyle\sum_{n=0}^\infty a(c+n\lambda), & 若\ \mu < \infty, \\ 0, & 若\ \mu = \infty. \end{cases}$$

该定理有另一叙述形式, 它等价于上述刚刚给出的形式, 但更为直接地利用了更新函数. 设 $h > 0$ 是已知的, 在方程 (5.3) 中特别指定

$$a(y) = \begin{cases} 1, & 若\ 0 \leqslant y < h, \\ 0, & 若\ h \leqslant y. \end{cases}$$

对于 $t > h$, 由于 (4.10), 故我们有

$$\begin{aligned} A(t) &= a(t) + \int_0^t a(t-x)\mathrm{d}M(x) \\ &= \int_{t-h}^t \mathrm{d}M(x) \\ &= M(t) - M(t-h), \end{aligned}$$

并且 $\mu^{-1}\int_0^\infty a(x)\mathrm{d}x = h/\mu$. 如果 F 不是算术的, 我们可根据更新定理推得

$$\lim_{t\to\infty}[M(t) - M(t-h)] = h/\mu, \tag{5.4}$$

这里约定当 $\mu = \infty$ 时, 有 $h/\mu = 0$.

反之, 用阶梯函数逼近直接黎曼可积, 定理 (5.1) 即可由式 (5.4) 推出. 更新定理的第二种形式的严格叙述如下:

定理 5.2　设 F 是均值为 μ 的正随机变量的分布函数. 令 $M(t)=\sum_{k=1}^{\infty}F_k(t)$ 是关于 F 的更新函数. 设 $h>0$ 是固定的.

(i) 如果 F 不是算术的, 则

$$\lim_{t\to\infty}[M(t+h)-M(t)]=h/\mu.$$

(ii) 如果 F 是算术的, 且 h 是跨距 λ 的倍数, 则上述极限同样成立.

在结束本节之前, 我们可以把定理 4.2, 即初等更新定理:

$$\lim_{t\to\infty}\frac{1}{t}M(t)=\frac{1}{\mu}, \tag{5.5}$$

作为定理 5.2 的推论重新推导证得. 为此, 置 $b_n=M(n+1)-M(n)$. 若设 F 不是算术的, 定理 5.2 指出当 $n\to\infty$ 时 $b_n\to 1/\mu$, 所以 b_n 的平均值也收敛于相同的极限. 这样,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}b_k=\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}[M(k+1)-M(k)]=\lim_{n\to\infty}\frac{1}{n}M(n)=\frac{1}{\mu}.$$

现在, 对任意 $t>0$, 令 $[t]$ 表示不超过 t 的最大整数. 由 $M(t)$ 的单调性知

$$\frac{[t]}{t}\frac{M([t])}{[t]}\leqslant\frac{M(t)}{t}\leqslant\frac{[t]+1}{t}\frac{M([t]+1)}{[t]+1}.$$

由于 $t^{-1}M(t)$ 被收敛于 μ^{-1} 的函数列所夹住, 从而推得式 (5.5). 如果 F 是算术的, 其跨距为 λ, 我们置 $b_n=M[(n+1)\lambda]-M(n\lambda)$, 然后利用与上述完全平行的推理, 即可证得 (ii).

5.6　更新定理的应用

(a) 剩余寿命的极限分布

设 $\gamma_t=S_{N(t)+1}-t$ 是在时刻 t 的剩余寿命, 固定 $z>0$, 置

192

$$A_z(t)=\Pr\{\gamma_t>z\}.$$

我们应用更新论证法构造关于 A_z 的更新方程. 依通常途径假定第一次更新时间为

$X_1 = x$, 得到 (建议读者画图分析)

$$\Pr\{\gamma_t > z | X_1 = x\} = \begin{cases} 1, & 若\ x > t + z, \\ 0, & 若\ t + z \geqslant x > t, \\ A_z(t - x), & 若\ t \geqslant x > 0. \end{cases}$$

再根据全概率公式

$$\begin{aligned} A_z(t) &= \int_0^\infty \Pr\{\gamma_t > z | X_1 = x\} \mathrm{d}F(x) \\ &= 1 - F(t + z) + \int_0^t A_z(t - x) \mathrm{d}F(x). \end{aligned} \tag{6.1}$$

由定理 4.1 推得

$$A_z(t) = 1 - F(t + z) + \int_0^t \{1 - F(t + z - x)\} \mathrm{d}M(x).$$

为了得到极限分布, 我们假设

$$\mu = E[X_1] = \int_0^\infty \{1 - F(x)\} \mathrm{d}x < \infty.$$

则

$$\int_0^\infty \{1 - F(t + z)\} \mathrm{d}t = \int_z^\infty \{1 - F(y)\} \mathrm{d}y < \infty,$$

并且由于 $\{1 - F(t + z)\}$ 是单调的, 从而当 z 固定时, 它作为 t 的函数是直接黎曼可积的. 应用更新定理得到

$$\lim_{t \to \infty} \Pr\{\gamma_t > z\} = \lim_{t \to \infty} A_z(t) = \mu^{-1} \int_z^\infty \{1 - F(y)\} \mathrm{d}y, \quad z > 0, \tag{6.2}$$

这即是剩余寿命的渐近分布.

现龄 δ_t 和全寿命 β_t 的极限分布可由 (6.2) 推出. 借助图 5-6 我们可直接证明下面事件的等价性

$$\{\gamma_t \geqslant x \text{ 和 } \delta_t \geqslant y\} \quad \text{当且仅当} \quad \{\gamma_{t-y} \geqslant x + y\}. \tag{6.3}$$

由此推得

$$\begin{aligned} \lim_{t \to 0} \Pr\{\delta_t \geqslant y, \gamma_t \geqslant x\} &= \lim_{t \to \infty} \Pr\{\gamma_{t-y} \geqslant x + y\} \\ &= \mu^{-1} \int_{x+y}^\infty \{1 - F(z)\} \mathrm{d}z, \end{aligned} \tag{6.4}$$

193

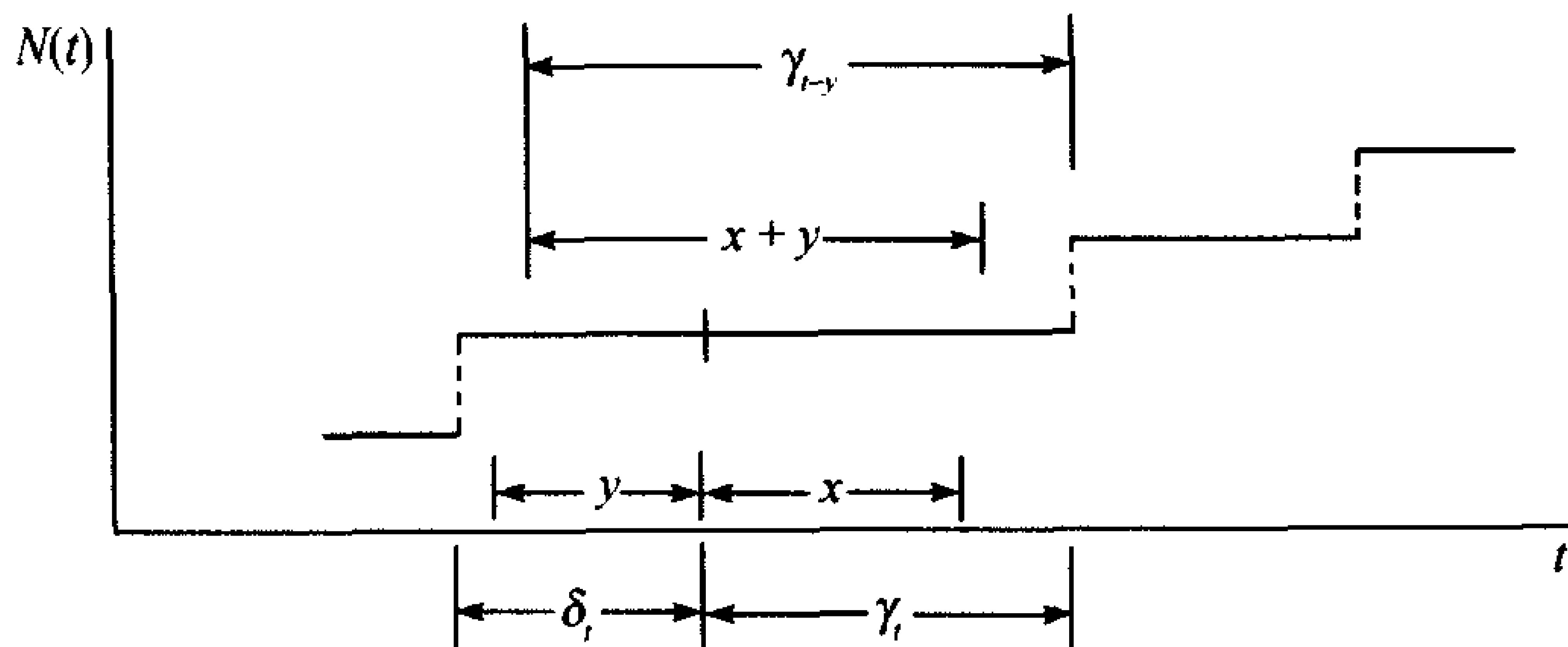


图 5-6　说明 $\{\delta_t \geqslant y$ 和 $\gamma_t \geqslant x\}$当且仅当 $\{\gamma_{t-y} \geqslant x+y\}$

式 (6.4) 即为 $\{\delta_t, \gamma_t\}$ 的极限联合分布. 特别地,

$$\begin{aligned}\lim_{t\to\infty} \Pr\{\delta_t \geqslant y\} &= \lim_{t\to\infty} \Pr\{\delta_t \geqslant y, \gamma_t \geqslant 0\} \\ &= \mu^{-1} \int_y^\infty \{1 - F(z)\}\mathrm{d}z. \qquad (6.5)\end{aligned}$$

$\beta_t = \delta_t + \gamma_t$ 的极限分布能够从极限公式 (6.4) 算出. 而且它可借助于更新论证法很快证明得到. 我们下面只作简单介绍. 定义

$$K_x(t) = \Pr\{\beta_t > x\}.$$

设定第一次更新事件的时间, 我们有

$$\Pr\{\beta_t > x | X_1 = y\} = \begin{cases} 1, & 若\ y > \max(x, t), \\ K_x(t-y), & 若\ y \leqslant t, \\ 0, & 其他. \end{cases}$$

由全概率公式推得更新方程

$$K_x(t) = 1 - F(\max(x,t)) + \int_0^t K_x(t-y)\mathrm{d}F(y).$$

应用更新定理得到如下极限分布

$$\begin{aligned}\lim_{t\to\infty} \Pr\{\beta_t > x\} = \lim_{t\to\infty} K_x(t) &= \frac{1}{\mu}\int_0^\infty [1 - F(\max(x,\tau))]\mathrm{d}\tau \\ &= \frac{1}{\mu}\int_x^\infty \xi \mathrm{d}F(\xi), \qquad (6.6)\end{aligned}$$

194

其中最后一个等式是对第一个积分使用分部求积分法得到的. 相应地, 我们构造全寿命的极限分布是

$$\lim_{t\to\infty} \Pr\{\beta_t \leqslant x\} = \frac{1}{\mu}\int_0^x \xi \mathrm{d}F(\xi) = G(x).$$

当 F 的密度为 f 时, 则 G 的密度显然是 $xf(x)/\mu$.

$G(x)$ 的均值与 $F(x)$ 的均值之间的关系是有趣的. 考虑

$$\int_0^\infty x\mathrm{d}G(x) = \frac{1}{\mu}\int_0^\infty x^2\mathrm{d}F(x), \tag{6.7}$$

把这个量与 μ 作比较, μ 是任意更新区间的平均长度. 注意, 施瓦兹不等式 (见第 1 章) 蕴涵着

$$\int_0^\infty x^2\mathrm{d}F(x) \geqslant \mu^2 = \Big(\int_0^\infty x\mathrm{d}F(x)\Big)^2, \tag{6.8}$$

除非 F 是退化分布, 否则严格不等号成立. 关系式 (6.8) 告诉我们, 目前而言平均全寿命一般严格超过普通平均寿命. 当然, 这个事实和泊松过程情况 (5.3 节) 计算结果是一致的, 后者指出了平均全寿命是平均寿命的两倍. 不等式 (6.8) 证实了对于包含指定点的寿命区间进行抽样具有 "长度偏倚" 性质.

(b) 更新函数的渐近展开式

假设 F 是非算术分布, 其方差 σ^2 有限. 现确定 $M(t)$ 渐近展开式中第二项, 需要证明

$$\lim_{t\to\infty}\{M(t) - \mu^{-1}t\} = \frac{\sigma^2 - \mu^2}{2\mu^2}.$$

当 t 很大时, $M(t)$ 的高阶渐近性质也可应用类似方法求得, 假设 F 的高阶矩存在.

定义

$$\begin{aligned} H(t) &= M(t) + 1 - \mu^{-1}t \\ &= E[N(t)+1] - \mu^{-1}t \\ &= \mu^{-1}\{E[S_{N(t)+1}] - t\} \quad (\text{由 } (4.12)) \\ &= \mu^{-1}E[\gamma_t] \quad (\text{由}(4.13)) \end{aligned}$$

195

再一次求助于更新论证法, 设定第一次更新时间 $X_1 = x$, 得到一个关于 $H(t)$ 的更新方程. 直接列举各种情况算出

$$E[\gamma_t|X_1 = x] = \begin{cases} x - t, & \text{若 } x \geqslant t, \\ \mu H(t-x), & \text{若 } x < t. \end{cases}$$

应用全概率公式得到

$$\begin{aligned} \mu H(t) &= \int_0^\infty E[\gamma_t|X_1 = x]\mathrm{d}F(x) \\ &= \int_t^\infty (x-t)\mathrm{d}F(x) + \mu\int_0^t H(t-x)\mathrm{d}F(x). \end{aligned}$$

由于

$$\int_t^{\infty}(x-t)\mathrm{d}F(x)=\int_0^{\infty}y\mathrm{d}F(t+y)=\int_0^{\infty}\{1-F(t+y)\}\mathrm{d}y$$

是 t 的单调函数, 利用表达式 $1-F(t+y)=\int_{t+y}^{\infty}\mathrm{d}F(z)$, 并调换积分顺序, 得

$$\begin{aligned}
&\int_0^{\infty}\left\{\int_0^{\infty}[1-F(t+y)]\mathrm{d}y\right\}\mathrm{d}t\\
&=\int_0^{\infty}\int_0^{\infty}\int_{t+y}^{\infty}\mathrm{d}F(z)\mathrm{d}y\mathrm{d}t\\
&=\int_0^{\infty}\int_t^{\infty}\left\{\int_0^{z-t}\mathrm{d}y\right\}\mathrm{d}F(z)\mathrm{d}t\\
&=\int_0^{\infty}\int_t^{\infty}(z-t)\mathrm{d}F(z)\mathrm{d}t\\
&=\int_0^{\infty}\int_0^{z}(z-t)\mathrm{d}t\mathrm{d}F(z)\\
&=\frac{1}{2}\int_0^{\infty}z^2\mathrm{d}F(z)=\frac{1}{2}(\sigma^2+\mu^2)<\infty.
\end{aligned}$$

196

所以更新定理蕴涵着

$$\lim_{t\to\infty}\mu H(t)=\mu^{-1}\frac{1}{2}(\sigma^2+\mu^2),$$

或

$$\lim_{t\to\infty}\{M(t)-\mu^{-1}t\}=\lim_{t\to\infty}\{H(t)-1\}=\frac{\sigma^2+\mu^2}{2\mu^2}-1=\frac{\sigma^2-\mu^2}{2\mu^2},$$

即为所证.

5.7 更新过程的推广

A. 延迟更新过程

我们继续假设 $\{X_k\}$ 是独立正随机变量序列, 但仅从第二项开始, 即 $X_2, X_3, \cdots$ 是同分布的, 其分布函数为 F, 而 X_1 可能有不同的分布函数 G. 这样的过程称为**延迟更新过程**. 它除了从初始时刻到第一次更新具有特别分布之外, 其他过程与普通更新过程完全相同.

产生延迟更新过程的一种途径是当 $t=0$ 时运行的成分, 不是从新的状态开始的. 例如, 假设原点取在一个普通更新过程开始以后的第 y 个单位时间处, 则该延迟过程在原点之后到第一次更新时间的分布是一个普通更新过程在时刻 y 处的剩余寿命的分布.

如前, 设 $S_0 = 0$ 和 $S_n = X_1 + \cdots + X_n$, 令 $N(t)$ 是至时刻 t 为止的更新数目, 我们必须区分延迟更新过程中更新数目的均值

$$M_D(t) = E[N(t)]$$

和关于分布函数 F 的更新函数

$$M(t) = \sum_{k=1}^{\infty} F_k(t).$$

设定第一次更新时间, 记

$$E[N(t)|X_1 = x] = \begin{cases} 0, & 若\ x > t, \\ 1 + M(t-x), & 若\ x \leqslant t. \end{cases}$$

197

利用全概率公式, 得到

$$\begin{aligned} M_D(t) &= \int_0^{\infty} E[N(t)|X_1 = x]\mathrm{d}G(x) \\ &= \int_0^t \{1 + M(t-x)\}\mathrm{d}G(x) \\ &= G(t) + \int_0^t M(t-x)\mathrm{d}G(x) \\ &= G(t) + \int_0^t G(t-x)\mathrm{d}M(x). \end{aligned} \tag{7.1}$$

显然, 等式 (7.1) 显示 $M_D(t)$ 可作为下面更新方程 (与 (4.9) 和 (4.10) 相比较) 的解:

$$M_D(t) = G(t) + \int_0^t M_D(t-x)\mathrm{d}F(x). \tag{7.2}$$

我们以下证明, 若 F 不是算术分布, $M_D(t)$ 服从更新定理. (对于 F 是算术分布情况, 可类似处理). 由 (7.1) 我们可知, 对任意 $t > 0$ 有

$$M_D(t) = G(t) + \int_0^t M(t-x)\mathrm{d}G(x),$$

特别, 当 $t > h$ 时,

$$M_D(t-h) = G(t-h) + \int_0^{t-h} M(t-h-x)\mathrm{d}G(x).$$

约定当 $x < 0$ 时 $M(x) = 0$, 上面两个等式之差为

$$M_D(t) - M_D(t-h) = G(t) - G(t-h) + \int_0^t \{M(t-x) - M(t-h-x)\}\mathrm{d}G(x).$$

为方便起见把上式积分分为两个：

$$\int_0^{t/2}\{M(t-x)-M(t-h-x)\}\mathrm{d}G(x)+\int_{t/2}^{t}\{M(t-x)-M(t-h-x)\}\mathrm{d}G(x).$$

由于 $\lim_{t\to\infty}\{M(t-x)-M(t-h-x)\}=h/\mu$, 第一个积分收敛于 h/μ, 而第二个收敛于 0, 因为 $\{M(t-x)-M(t-h-x)\}$ 是收敛的, 因此是 x 的有界函数. 显然, 当 $t\to\infty$ 时, 有 $G(t)-G(t/2)\to 0$, 于是求和得

198

$$\lim_{t\to\infty}[M_D(t)-M_D(t-h)]=h/\mu.$$

B. 平稳更新过程

一个延迟更新过程, 当其第一个寿命分布函数为

$$G(x)=\mu^{-1}\int_0^x\{1-F(y)\}\mathrm{d}y$$

则称它为平稳更新过程. 我们想象一个更新过程开始于无穷远的过去, 使得正在原点运行的个体的剩余寿命具有普通更新过程剩余寿命的极限分布. 我们用 G 表示这个极限分布.

可以预期这样一个过程将呈现出若干平稳性或时间不变性. 对于平稳更新过程, 我们将对此作充分说明, 对所有 t,

$$M_D(t)=E[N(t)]\equiv t/\mu \tag{7.3}$$

和

$$\Pr\{\gamma_t^D\leqslant x\}=G(x)$$

均成立. 因此一般来说仅是渐近的更新关系式, 在平稳更新过程中, 却变成对于所有 t 成立的恒等式.

由等式 (7.2) 知 $M_D(t)$ 满足更新方程

$$M_D(t)=G(t)+\int_0^t M_D(t-x)\mathrm{d}F(x), \tag{7.4}$$

由于这样一个方程的解是唯一的 (在适当的有界性限制之下, 见定理 4.1), 我们仅需要检验 $M_D(t)\equiv t/\mu$ 是否满足 (7.4). 我们有

$$\begin{aligned}G(t)+\int_0^t M_D(t-x)\mathrm{d}F(x)&=\mu^{-1}\int_0^t\{1-F(x)\}\mathrm{d}x+\mu^{-1}\int_0^t(t-x)\mathrm{d}F(x)\\&=\mu^{-1}t+\mu^{-1}\left\{\int_0^t(t-x)\mathrm{d}F(x)-\int_0^t F(y)\mathrm{d}y\right\}\\&=\mu^{-1}t,\end{aligned}$$

199

通过分部积分可验证大括号里的式子为 0, 由此推得 $M_D(t) \equiv t/\mu$ 成立.

按同样方法可证, 对任意 x, $\Pr\{\gamma_t^D \leqslant x\} = G(x)$ 成立, 这里 γ_t^D 是延迟 (平稳) 更新过程中的剩余寿命. 令

$$A_x^D(t) = \Pr\{\gamma_t^D > x\} \quad 和 \quad A_x(t) = \Pr\{\gamma_t > x\},$$

其中 γ_t 是普通更新过程的剩余寿命. 由标准更新论证方法得到

$$A_x^D(t) = 1 - G(t+x) + \int_0^t A_x(t-y)\mathrm{d}G(y),$$

或

$$A_x^D(t) = 1 - G(t+x) + G * A_x(t). \tag{7.5}$$

其中更新方程

$$A_x(t) = 1 - F(t+x) + F * A_x(t)$$

早已在 5.6 节中作了详细介绍, 借助于定理 4.1, 其解可表示为形式

$$A_x(t) = a_x(t) + M * a_x(t), \tag{7.6}$$

其中

$$a_x(t) = 1 - F(t+x).$$

把 (7.6) 代入 (7.5), 并利用公式 $M_D(t) = G(t) + G * M(t)$(此式可直接由定义得到), 我们得到

$$\begin{aligned}
A_x^D(t) &= 1 - G(t+x) + G * a_x(t) + G * M * a_x(t) \\
&= 1 - G(t+x) + M_D * a_x(t) \\
&= 1 - G(t+x) + \int_0^t a_x(t-y)\mathrm{d}M_D(y).
\end{aligned}$$

由于 $a_x(t-y) = 1 - F(t+x-y)$ 和 $M_D(y) \equiv y/\mu$, 所以 $\mathrm{d}M_D(y) = \mu^{-1}\mathrm{d}y$. 则

$$\begin{aligned}
A_x^D(t) &= 1 - G(t+x) + \mu^{-1}\int_0^t \{1 - F(t+x-y)\}\mathrm{d}y \\
&= 1 - G(t+x) + \mu^{-1}\int_x^{t+x} \{1 - F(u)\}\mathrm{d}u \\
&= 1 - G(t+x) + G(t+x) - G(x) \\
&= 1 - G(x),
\end{aligned}$$

即为所证.

200

C. 累积的和相关的过程

假设与第 i 个元件或寿命区间相联系的, 除了寿命区间 X_i 之外还有第二个随机变量 $Y_i(\{Y_i\}$ 同分布). 我们允许 X_i 和 Y_i 是相关的, 但假设 $(X_1,Y_1),(X_2,Y_2),\cdots$ 是独立的. 我们将使用记号 $F(x)=\Pr\{X_i\leqslant x\}, G(y)=\Pr\{Y_i\leqslant y\}, \mu=E[X_i]$ 和 $\nu=E[Y_i]$.

实际上和理论上的若干问题可用如下术语加以描述.

I. 在每个更新区间含有两个分量的更新过程

假设 Y_i 表示 X_i 的一部分. 图 5-7 解释了这个模型. 图中 Y 部分出现在区间的开端, 但这并非下面结果的本质所在.

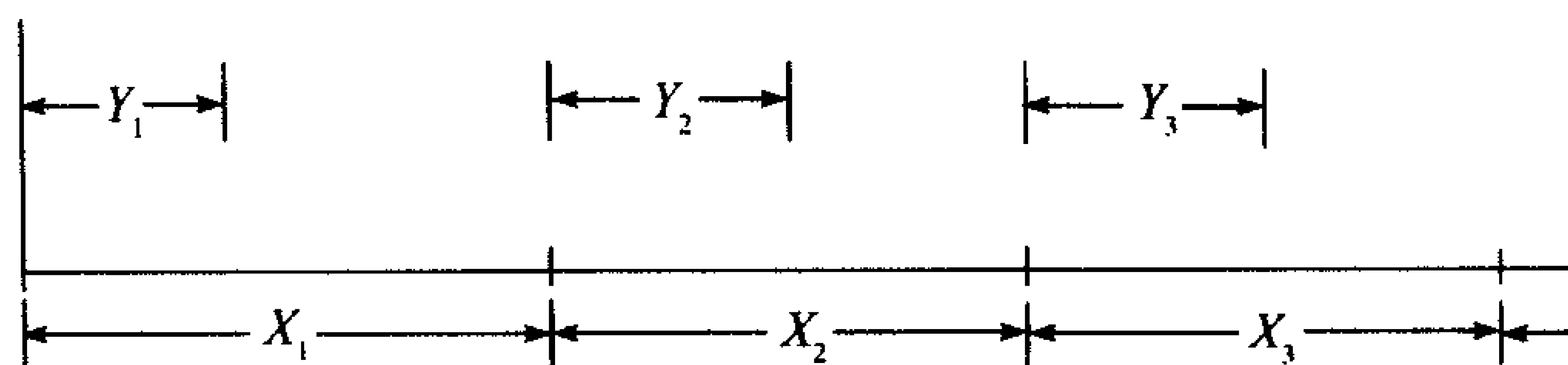


图 5-7 每个更新区间会有相应随机变量 Y_i 的更新过程

设 $p(t)$ 是 t 所在某个更新区间 Y 部分的概率. 设定第一次更新区间的长度 $X_1=x$, 并将其区别 $x<t$ 和 $x\geqslant t$ 两种可能, 我们得到 (利用通常方法) 更新方程

$$p(t)=\Pr\{t \text{ 被 } Y_1 \text{ 所覆盖}\}+\int_0^t p(t-\xi)\mathrm{d}F(\xi).$$

令

201

$$I_{Y_1}(t)=\begin{cases}1, & \text{若 } Y_1 \text{ 覆盖 } t,\\ 0, & \text{若 } Y_1 \text{ 不覆盖 } t.\end{cases}$$

则

$$\Pr\{t \text{ 被 } Y_1 \text{ 所覆盖}\}=E[I_{Y_1}(t)],$$

且

$$\begin{aligned}\int_0^\infty \Pr\{t \text{ 被 } Y_1 \text{ 所覆盖}\}\mathrm{d}t &= \int_0^\infty E[I_{Y_1}(t)]\mathrm{d}t\\ &=E\left[\int_0^\infty I_{Y_1}(t)\mathrm{d}t\right]\\ &=E[Y_1]=\nu,\end{aligned}$$

由于被 Y_1 所覆盖的点的全体仍为 Y_1, 故应用更新定理, 我们可断言, 如果 F 是非

算术的, 并且 Pr{t 被 Y_1 所覆盖}是直接黎曼可积的, 则

$$\lim_{t\to\infty} p(t) = \mu^{-1}\int_0^\infty \Pr\{t \text{ 被 } Y_1 \text{ 所覆盖}\}\mathrm{d}t \\ = \nu/\mu. \tag{7.7}$$

下面是一些具体例子.

(a) 替换模型

考虑替换不是瞬间进行的替换模型的情况. 设 Y_i 是第 i 个元件运行的时间, Z_i 是第 $i+1$ 个元件运行前的滞后区间. (在替换中的滞后可以解释为服务元件修理的时间). 我们假设相继替换之间的时间序列为 $X_k = Y_k + Z_k, k = 1, 2, \cdots$, 由此构成一个更新过程. 若 X_i 的分布是非算术的, 则在时刻 t 系统处于运行状态的概率 $p(t)$ 收敛于 $E[Y_1]/E[X_1]$.

(b) 排队模型

若顾客的到达遵循泊松过程, 那么从第 k 个忙碌周期至下一个忙碌周期的间隔时间构成一个更新过程. (一个忙碌周期是排队队伍不为空的持续时间). 每个 X_k 是由忙碌部分 Z_k 和空闲部分 Y_k 组成. 那么, 队伍在时刻 t 是空的概率 $p(t)$ 收敛于 $E[Y_1]/E[X_1]$.

(c) 计数问题

设 $X_k, k = 1, 2, \cdots$, 表示计数器相继记录两个粒子的间隔时间序列, Y_k 表示在 X_k 更新区间里的闭锁时间, 则计数器在时刻 t 闭锁的概率 $p(t)$ 收敛于 $E(Y_1)/E(X_1)$. 202

II. 累积过程

把 Y_i 解释为与第 i 个更新周期相联系的成本或价值等. 现在我们来考虑一般的关于一对随机变量 (X_i, Y_i) 问题, 其中 X_i 产生一个更新过程. 这里感兴趣的主要在于所谓的**累积过程**:

$$W(t) = \sum_{k=1}^{N(t)+1} Y_k,$$

它表示到时刻 t 为止累积的价值等 (假定交易在每个更新周期开始进行的). 设定第一次更新时间 $X_1 = x$, 并依 $x > t$ 和 $x \leqslant t$ 两种可能情况, 我们获得对于 $A(t) = E[W(t)]$ 的更新方程

$$A(t) = E[Y_1] + \int_0^t A(t-x)\mathrm{d}F(x).$$

根据定理 4.1 得

$$A(t) = E[Y_1] + \int_0^t E[Y_1]\mathrm{d}M(x) = E[Y_1][1 + M(t)].$$

由此, 当 F 是非算术的且 $h > 0$ 时, 有

$$\lim_{t\to\infty} [A(t) - A(t-h)] = E[Y_1]h/\mu,$$

并在任何情况下都成立下式

$$\lim_{t\to\infty} \frac{1}{t}A(t) = E[Y_1]/\mu.$$

这就验证了 $E[Y_1]/\mu$ 可作为每单位时间长久平均成本、价值等的解释, 这在 5.3 节例子中曾多次被使用.

下面是累积过程的一些例子.

(a) 替换模型

在依龄替换策略之下 (见 5.3 节例 B), 我们假设在时刻 T 替换的成本是 c_1 元, 而在时刻 $x < T$ 损坏替换的成本是 c_2 元. 若 Y_k 是第 k 次替换周期所花费的成本, 则

203

$$Y_k = \begin{cases} c_1, & \text{以概率 } 1 - F(T), \\ c_2, & \text{以概率 } F(T), \end{cases}$$

因此 $E[Y_k] = c_1[1 - F(T)] + c_2F(T)$. 由于一个替换周期的平均长度是

$$E[\min\{X_k, T\}] = \int_0^T [1 - F(x)]\mathrm{d}x,$$

故每单位时间长久成本是

$$\frac{c_1[1 - F(T)] + c_2F(T)}{\displaystyle\int_0^T [1 - F(x)]\mathrm{d}x},$$

在一些特殊情况下, 可借助微积分方法或数值计算方法求出使上式为最小的 T 值.

在整体替换策略中, 每 T 个单位时间有一次计划替换和平均 $M(T)$ 次损坏替换. 这样成本均值为 $E[Y_k] = c_1 + c_2M(T)$, 因而每单位时间长久平均成本是 $\{c_1 + c_2M(T)\}/T$.

(b) 计数模型

在计数器模型中 (见 5.3 节例 C), 设 Y_k 是第 $k-1$ 次记录信号和第 k 次记录信号之间出现的但未被记录的信号的数目. 则每单位时间未被记录的粒子的长久平均数是 $E[Y_1]/E[X_1]$.

(c) **保险理论**

假设一个保险公司收到索取赔偿费要求的间隔时间序列 $\{X_i\}$ 为一更新过程. Y_k 是第 k 次索取的数量, 则 $W(t)=\sum_{k=0}^{N(t)+1} Y_k$ 表示至时刻 t 为止累积索取量, 长久平均索取率为

$$\lim_{t\to\infty}\frac{1}{t}E[W(t)]=E[Y_1]/E[X_1].$$

D. 有终止的更新过程

设一个更新过程可能有无限长的间隔时间, 这样的过程称为有终止的更新过程, 因为此过程在第一个无限长的间隔时间里停止更新. 这种情形如图 5-8 所示. 204

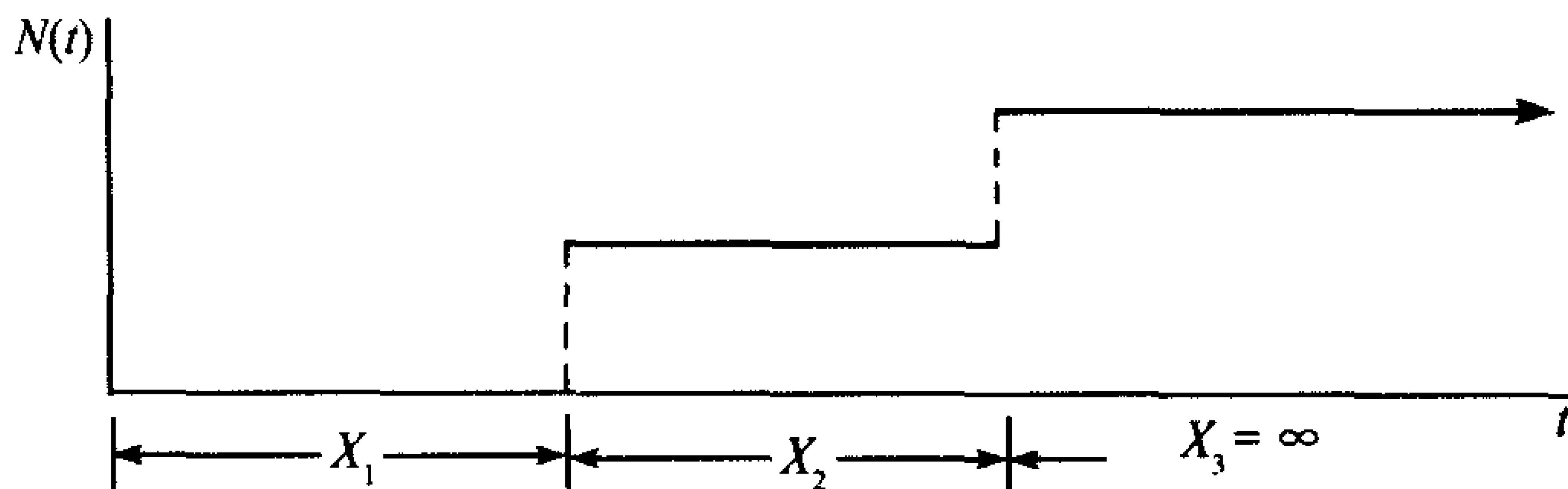


图 5-8

令 $L=F(\infty)=\Pr\{X_k<\infty\}<1$ 和 $1-L=\Pr\{X_k=\infty\}>0$. 在全部时间里更新的总数用 $N(\infty)$ 表示, 则它是有限值随机变量并服从几何概率分布

$$\Pr\{N(\infty)\geqslant k\}=L^k,\quad k=0,1,2,\cdots,$$

因此

$$\begin{aligned}E[N(\infty)]&=\sum_{k=1}^{\infty}\Pr\{N(\infty)\geqslant k\}\\&=L/(1-L).\end{aligned}$$

有终止更新过程的实现仍具有性质: $N(t)\geqslant k$ 当且仅当 $S_k\leqslant t$, 所以

$$\Pr\{N(t)\geqslant k\}=\Pr\{S_k\leqslant t\}=F_k(t),$$

并且

$$M(t)=E[N(t)]=\sum_{k=1}^{\infty}F_k(t)<\sum_{k=1}^{\infty}L^k=L/(1-L).$$

再者, 利用更新论证法, 得到方程

$$M(t)=F(t)+\int_0^t M(t-x)\mathrm{d}F(x).$$

然而, 由于 F 不是正常的概率分布函数, 更新定理并不能直接使用. 幸运的是, 有一常见方法可解决这个问题. 假设

$$g(s)=\int_0^{\infty}\mathrm{e}^{sx}\mathrm{d}F(x)$$

当 $s\geqslant 0$ 时是 s 的有限函数. 则 g 将是连续的并且 $g(0)=L<1$, $\lim\limits_{s\to\infty}g(s)=\infty$, 从而推知存在一个唯一的正值 $S_0=\lambda>0$, 使得

205

$$g(\lambda)=\int_0^{\infty}\mathrm{e}^{\lambda x}\mathrm{d}F(x)=1.$$

定义

$$\hat{F}(t)=\int_0^{t}\mathrm{e}^{\lambda x}\mathrm{d}F(x),$$

则 $\hat{F}(t)$ 是不减的并且 $\lim\limits_{t\to\infty}\hat{F}(t)=1$, 因此 $\hat{F}$ 是正常的分布函数. 现在我们考虑更新方程

$$A(t)=a(t)+\int_0^{t}A(t-x)\mathrm{d}F(x).$$

置 $\hat{A}(t)=\mathrm{e}^{\lambda t}A(t),\hat{a}(t)=\mathrm{e}^{\lambda t}a(t)$, 则有

$$\begin{aligned}\hat{A}(t)&=\mathrm{e}^{\lambda t}A(t)\\&=\mathrm{e}^{\lambda t}a(t)+\int_0^{t}\mathrm{e}^{\lambda(t-x)}A(t-x)\mathrm{e}^{\lambda x}\mathrm{d}F(x)\\&=\hat{a}(t)+\int_0^{t}\hat{A}(t-x)\mathrm{d}\hat{F}(x),\end{aligned}$$

上式指出 $\hat{A}$ 满足包含正常分布函数的更新方程, 可对上述更新方程应用更新定理. 作为特例, 考虑 $A(t)=M(\infty)-M(t)=L/(1-L)-M(t)$. 等价地, $A(t)=E[N(\infty)-N(t)]$ 是满足关系式 $t<S_n<\infty$ 的指标 n 的平均数目. 我们现在来推导 $A(t)$ 所满足的更新方程. 事实上, 我们有

$$E[N(\infty)-N(t)|X_1=x]=\begin{cases}1+L/(1-L), & \text{若 } x>t,\\ A(t-x), & \text{若 } 0<x\leqslant t.\end{cases}$$

利用全概率公式推得

$$\begin{aligned}A(t)&=\int_0^{\infty}E[N(\infty)-N(t)|X_1=x]\mathrm{d}F(x)\\&=\{L-F(t)\}/(1-L)+\int_0^{t}A(t-x)\mathrm{d}F(x).\end{aligned}$$

其次验证

$$\hat{a}(t) = e^{\lambda t}\{L - F(t)\}/(1 - L)$$

是直接黎曼可积. 因为

$$\begin{aligned}\int_0^\infty \hat{a}(t)dt &= \frac{1}{1-L}\int_0^\infty e^{\lambda t}\{L - F(t)\}dt \\ &= \frac{1}{1-L}\int_0^\infty e^{\lambda t}\int_t^\infty dF(x)dt \\ &= \frac{1}{1-L}\int_0^\infty \int_0^x e^{\lambda t}dtdF(x) \\ &= \frac{1}{1-L}\int_0^\infty \frac{(e^{\lambda x}-1)}{\lambda}dF(x) \\ &= \frac{1}{1-L}\frac{(1-L)}{\lambda} = \frac{1}{\lambda}.\end{aligned}$$

206

故

$$\begin{aligned}\lim_{t\to\infty}\hat{A}(t) &= \lim_{t\to\infty} e^{\lambda t}[M(\infty) - M(t)] \\ &= \left\{\lambda\int_0^\infty x e^{\lambda x}dF(x)\right\}^{-1}.\end{aligned}$$

我们得出, $M(t)$ 以 λ 为比率的指数速度逼近 $M(\infty) = L/(1-L)$.

E. 交替和马尔可夫更新过程

一个交替更新过程是一个独立随机变量序列 $Y_1, Y_2, \cdots$, 其中

$$\begin{aligned}&Y_1, Y_{r+1}, Y_{2r+1}, \cdots \text{ 有分布函数 } F_1, \\ &Y_2, Y_{r+2}, Y_{2r+2}, \cdots \text{ 有分布函数 } F_2, \\ &\vdots \\ &Y_r, Y_{2r}, Y_{3r}, \cdots \text{ 有分布函数 } F_r.\end{aligned}$$

我们考虑一个相继通过状态 $1, 2, \cdots, r, 1, 2, \cdots, r, 1, 2, \cdots$ 的系统. 当经过每一个状态时都逗留一个随机时间段.

设 $P_i(t)$ 是系统在时刻 t 处于状态 i 的概率. 由 C 部分中的关系式 (7.7), 倘若分布 $F = F_1 * F_2 * \cdots * F_r$ 是非算术的, 我们推得

$$\lim_{t\to\infty} p_i(t) = \mu_i/(\mu_1 + \cdots + \mu_r),$$

其中

$$\mu_i = E[Y_i] < \infty, \quad i = 1, \cdots, r.$$

根据具有转移概率矩阵 $\boldsymbol{P}=\|P_{ij}\|_{i,j=1}^{r}$ 的马尔可夫链知识, 马尔可夫更新或半马尔可夫过程经过状态 $1,\cdots,r$ 并且已知下一状态是 j, 则逗留在 i 的时间具有分布函数 F_{ij}, 且限于这一串状态, 所有逗留时间都可假定是独立的. 在状态 i 逗留时间的无条件分布函数是 $F_i(t)=\sum_{j=1}^{r}P_{ij}F_{ij}(t)$, 其中假设它具有有限均值 μ_i. 此外, 还假设上面马尔可夫链是不可约的和常返的, 并具有已知的平稳分布 $\pi_j=\sum_i \pi_i P_{ij}$.

207

现设过程从固定状态 i 开始, 并令 k 是事先指定的状态. 从某一次到达状态 i 到下一次到达状态 i 之间的时间区间称为一个 i 循环. 这些相继到达 i 的间隔时间序列构成一个更新过程. 假设至少一个 F_i 不是算术的, 由关系式 (7.7) 可知, 在 t 时刻处于状态 k 的概率 $P_k(t)$ 收敛于在状态 k 的平均逗留时间除以 i 循环的平均时间长度所得的数值. 再由全概率公式, 一个 i 循环中处于状态 k 的平均时间等于 μ_k 乘以在相继两次到达状态 i 的间隔时间中到达 k 的平均次数. 第二个因子仅依赖于逗留状态的离散时间马尔可夫链, 因此必须与 π_k 成比例关系. 由此推得

$$\lim_{t\to\infty}p_k(t)=c\pi_k\mu_k,$$

其中 c 是比例常数. 由于这些概率和必为 1, 因此 $c=1/(\pi_1\mu_1+\cdots+\pi_r\mu_r)$.

F. 更新过程的中心极限定理

定理 7.1 设 $\{X_n\}$ 是一更新过程, $\mu=E[X_1]<\infty$, 并且 $\sigma^2=E[(X_1-\mu)^2]<\infty$, 则

$$\lim_{t\to\infty}\Pr\left\{\frac{N(t)-t/\mu}{\sqrt{t\sigma^2/\mu^3}}<x\right\}=\Phi(x),$$

其中

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left(-\frac{1}{2}u^2\right)\mathrm{d}u$$

是正态积分.

证明 证明是基于对 $S_n=X_1+X_2+\cdots+X_n$ 使用中心极限定理及基本等价关系: $\{N(t)<n\}$ 当且仅当 $\{S_n>t\}$.

设 x 是固定的, 并令 n,t 以使下面关系式成立的方式趋于 ∞:

208

$$\lim_{\substack{t\to\infty\\ n\to\infty}}\frac{t-n\mu}{\sigma\sqrt{n}}=-x.$$

则由通常的中心极限定理有

$$\lim_{\substack{t\to\infty\\ n\to\infty}}\Pr\{S_n>t\}=\lim_{\substack{t\to\infty\\ n\to\infty}}\Pr\left\{\frac{S_n-n\mu}{\sigma\sqrt{n}}>-x\right\}=1-\Phi(-x)=\Phi(x).$$

然而, 另一方面, 因为当 $t\to\infty, n\to\infty$ 并使得 $(t-n\mu)/\sqrt{n\sigma^2}\to -x$ 成立时, 有 $(n-t/\mu)/\sqrt{t\sigma^2/\mu^3}\to x$, 故

$$\begin{aligned}\Phi(x) &= \lim_{\substack{t\to\infty\\ n\to\infty}} \Pr\{S_n>t\}\\ &= \lim_{\substack{t\to\infty\\ n\to\infty}} \Pr\{N(t)<n\}\\ &= \lim_{\substack{t\to\infty\\ n\to\infty}} \Pr\left\{\frac{N(t)-t/\mu}{\sqrt{t\sigma^2/\mu^3}}<\frac{n-t/\mu}{\sqrt{t\sigma^2/\mu^3}}\right\}\\ &= \lim_{t\to\infty} \Pr\left\{\frac{N(t)-t/\mu}{\sqrt{t\sigma^2/\mu^3}}<x\right\}.\end{aligned}$$

■

上面的分析只是形式的推导, 读者可以将其严格化.

G. 保险理论中的破产

令 $N(t)$ 表示在时间区间 $(0,t]$ 内向一个保险公司索取赔偿费的次数. 设 $N(t)$ 是参数为 λ 的泊松过程. 此外, 假设相继索取的金额 $Y_1, Y_2\cdots$ 是具有分布函数 $G(x)$ 的独立同分布随机变量. 令每单位时间的资金流入 (保险费, 投资等) 为 c 元, 并假设公司初始的财富为 z. 那么, 在时刻 t 现金余额应为

$$\Gamma(t)=z+ct-\sum_{i=1}^{N(t)}Y_i,$$

其中 Y_i 是第 i 次索取的金额. 下面一个有趣的问题是弄清楚具有连续偿付能力的概率, 它是 z 的函数. 即我们希望确定

$$\begin{aligned}R(z) &= \Pr\left\{z+ct-\sum_{i=1}^{N(t)}Y_i>0,\ \text{对所有 } t\right\} \qquad (7.8)\\ &= \text{在初始财富为 } z \text{ 下不破产的概率.}\end{aligned}$$

209

我们应用更新论证法, 设定第一次泊松事件的时间为 T_1, 利用全概率公式得到

$$R(z)=\int_0^\infty \Pr\left\{z+ct-\sum_{i=1}^{N(t)}Y_i>0 \text{ 对所有 } t|T_1=\tau\right\}\lambda e^{-\lambda\tau}d\tau. \qquad (7.9)$$

而若设定 Y_1 的值又可得到

$$\begin{aligned}&\Pr\left\{z+ct-\sum_{i=1}^{N(t)}Y_i>0 \text{ 对所有 } t|T_1=\tau\right\}\\ &=\int_0^\infty \Pr\left\{z+ct-\sum_{i=1}^{N(t)}Y_i>0 \text{ 对所有 } t|Y_1=y, T_1=\tau\right\}dG(y). \qquad (7.10)\end{aligned}$$

给定 $T_1=\tau, Y_1=y$, 过程 $\Gamma(t)$ 在时刻 τ 之后立刻更新, 拥有新的财富 $z+c\tau-y$. 所以

$$\Pr\left\{z+ct-\sum_{i=1}^{N(t)}Y_i>0 \text{ 对所有 } t|Y_1=y, T_1=\tau\right\}=R(z+c\tau-y). \tag{7.11}$$

显然, 当 $u<0$ 时, 有 $R(u)=0$ 成立.

把 (7.10) 和 (7.11) 代入 (7.9) 可得积分方程

$$R(z)=\int_0^{\infty}\Big(\int_0^{z+c\tau}R(z+c\tau-y)\mathrm{d}G(y)\Big)\lambda\mathrm{e}^{-\lambda\tau}\mathrm{d}\tau.$$

对外面的积分进行变量替换：$t=z+c\tau$, 并移项得

$$R(z)\mathrm{e}^{-\lambda z/c}=\frac{\lambda}{c}\int_z^{\infty}\Big(\int_0^t R(t-y)\mathrm{d}G(y)\Big)\mathrm{e}^{-\lambda t/c}\mathrm{d}t.$$

由此可知, $R(z)$ 是可微的, 微分后得到

$$\mathrm{e}^{-\lambda z/c}\Big[R'(z)-\frac{\lambda}{c}R(z)\Big]=-\frac{\lambda}{c}\mathrm{e}^{-\lambda z/c}\int_0^z R(z-y)\mathrm{d}G(y),$$

或等价地

210

$$R'(z)=\frac{\lambda}{c}R(z)-\frac{\lambda}{c}\int_0^z R(z-y)\mathrm{d}G(y).$$

对上式两端关于 z 积分得

$$R(w)-R(0)=\frac{\lambda}{c}\int_0^w R(z)\mathrm{d}z-\frac{\lambda}{c}\int_0^w\Big(\int_0^z R(z-y)\mathrm{d}G(y)\Big)\mathrm{d}z.$$

改变积分顺序, 并进行变量替换 $\xi=z-y$, 导出

$$R(w)=R(0)+\frac{\lambda}{c}\int_0^w R(z)\mathrm{d}z-\frac{\lambda}{c}\int_0^w\Big(\int_0^{w-y}R(\xi)\mathrm{d}(\xi)\Big)\mathrm{d}G(y).$$

定义 $S(x)=\int_0^x R(\xi)\mathrm{d}\xi$, 并利用分部求积分法得到

$$R(w)-R(0)=\frac{\lambda}{c}S(w)-\frac{\lambda}{c}\Big\{S(w)-\int_0^w R(w-y)[1-G(y)]\mathrm{d}y\Big\},$$

或

$$R(w)=R(0)+\frac{\lambda}{c}\int_0^w R(w-y)[1-G(y)]\mathrm{d}y.$$

从而, 我们得到了一个带有非正常密度函数 $(\lambda/c)[1-G(y)]$ 的更新方程, 因为

$$\int_0^\infty \frac{\lambda}{c}[1-G(y)]\mathrm{d}y = \frac{\lambda}{c}E[Y_1] = \frac{\lambda}{c}\mu.$$

如果 $\lambda\mu/c > 1$, 可以肯定 $R(z) = 0$(为什么?). (注意 $\lambda\mu$ 是每单位时间因索取赔偿而付出资金的期望值, 而 c 是流入率). 下面我们假定 $\lambda\mu/c < 1$. 由于 $a(w) = R(0)$, 在这个退化情形下仍应用定理 4.1, 可得

$$R(w) = a(w) + \int_0^w a(w-y)\mathrm{d}M(y) = R(0)[1+M(w)],$$

由于 $M(w)$ 对应于有终止的更新过程

$$\lim_{w\to\infty} M(w) = L/(1-L) = \frac{\lambda\mu/c}{1-\lambda\mu/c},$$

因此有

$$\lim_{w\to\infty} R(w) = \frac{R(0)}{1-(\lambda\mu/c)}.$$

211

但 $R(\infty) = 1$(为什么?), 从而得到

$$R(0) = 1 - \frac{\lambda\mu}{c}.$$

细致的分析将能得到更精确的渐近关系式.

5.8 更新理论更复杂的应用

A. 有变异的遗传模型

考虑一个总数为 N 的有限群体, 并用 $1, 2, \cdots, N$ 编号. 设它们是进化过程中第一代, 此过程受一定的自然选择作用和变异作用的影响.

群体的每一个体都具有称之为 "适合度" 的特征. 粗略地说, "适合度" 是个体在繁殖后代方面相对优势的一种衡量. 用 w_k^1 表示第一代的第 k 个个体的适合度. 取

$$u_k = \frac{w_k^1}{\sum_{j=1}^N w_j^1}, \quad k = 1, 2, \cdots, N, \tag{8.1}$$

它表示第 k 个个体的相对适合度的值, 下一代由 N 次独立抽样来形成, 每次抽样均服从概率矢量 (8.1) 的多项式分布. 于是, 每个后代都带有其亲本型的适合度值 w_k^1,

并且被挑选出来的概率是 u_k. 显然, 适合度比其他高的个体相应地具有较大的相对适合度值, 因此显然有比较大的机会繁殖其嫡亲的类型. 这个多项式繁殖过程产生的后代群体具有适合度值

$$\widetilde{\boldsymbol{w}}^2 = (\widetilde{w}_1^2, \widetilde{w}_2^2, \cdots, \widetilde{w}_N^2). \tag{8.2}$$

(自然, 每个 $\widetilde{w}_i^2$ 都是 $\{w_k^1\}$ 中的一个值. 个体的类型和它的适合度相一致.)

向量 $\widetilde{\boldsymbol{w}}^2$ 还没有形成成熟的第二代群体. 我们将引入变异的可能性, 后代可

212

能要经历适合度值的自发性的改变. 关于变异作用的精确叙述如下: 设$\{V_i^j; i = 1, 2, \cdots, N, j = 2, \cdots\}$ 是独立同分布正随机变量矩阵列, 令

$$\begin{aligned} w_1^2 &= \widetilde{w}_1^2 V_1^2, \\ w_2^2 &= \widetilde{w}_2^2 V_2^2, \\ &\vdots \\ w_N^2 &= \widetilde{w}_N^2 V_N^2. \end{aligned}$$

向量 $(w_1^2, w_2^2, \cdots, w_N^2)$ 表示第二代成熟个体的适合度. 重复上述过程, 相继产生一系列 N 维向量, 它们通过适合度向量 $\boldsymbol{w}^k = (w_1^k, \cdots, w_N^k)$ 及相对适合度向量 $\boldsymbol{u}^k = (u_1^k, \cdots, u_N^k)$[上标表示世代序数, 它从 (8.1) 所指定的初始群体开始计数] 的改变来描述群体的进化. 从 $\boldsymbol{w}^k$ 和 $\boldsymbol{u}^k$ 确定 $\boldsymbol{w}^{k+1}$ 所服从的概率规则与 $\boldsymbol{w}^1$ 形成 $\boldsymbol{w}^2$ 的规律一致, 我们叙述如下: 挑选 $w_i^k (i = 1, 2, \cdots N)$ 的概率为 u_i^k 以相同的概率从 $w_1^k, w_2^k, \cdots, w_N^k$ 中抽取 N 个独立的值. $\widetilde{\boldsymbol{w}}^{k+1} = (\widetilde{w}_1^{k+1}, \cdots, \widetilde{w}_N^{k+1})$ 表示所得的向量. 变异变化则通过乘以正随机变量 V_i^{k+1} 将 $\widetilde{\boldsymbol{w}}^{k+1}$ 变换成 $\boldsymbol{w}^{k+1}$, 即

$$w_i^{k+1} = \widetilde{w}_i^{k+1} V_i^{k+1}, \quad i = 1, 2, \cdots, N.$$

最后, 按如下规则确定相对适合度向量 $\boldsymbol{u}^{k+1}$

$$u_i^{k+1} = \frac{w_i^{k+1}}{\sum\limits_{j=1}^{N} w_j^{k+1}}, \quad i = 1, 2, \cdots, N.$$

这个进化过程可以用点 $\boldsymbol{w}^k (k = 1, 2, \cdots)$ 在 N 维空间所通过的路径来理解.

相对适合度 $\boldsymbol{u}^k$ 是随机向量 $\boldsymbol{w}^k$ 在 N 维单纯形

$$\Delta_N = \{x = (x_1, \cdots, x_N) : x_i \geqslant 0 \text{ 且 } x_1 + \cdots + x_N = 1\}$$

上的投影. 图 5-9 阐述了 $N = 3$ 时投影情况. 随着群体一代传一代, $\boldsymbol{u}^k$ 的变化可以用相对适合度点在单纯形 Δ_N 上的运动来表示. 自然, 我们需要弄清 $\boldsymbol{u}^k$ 的长久统计性态.

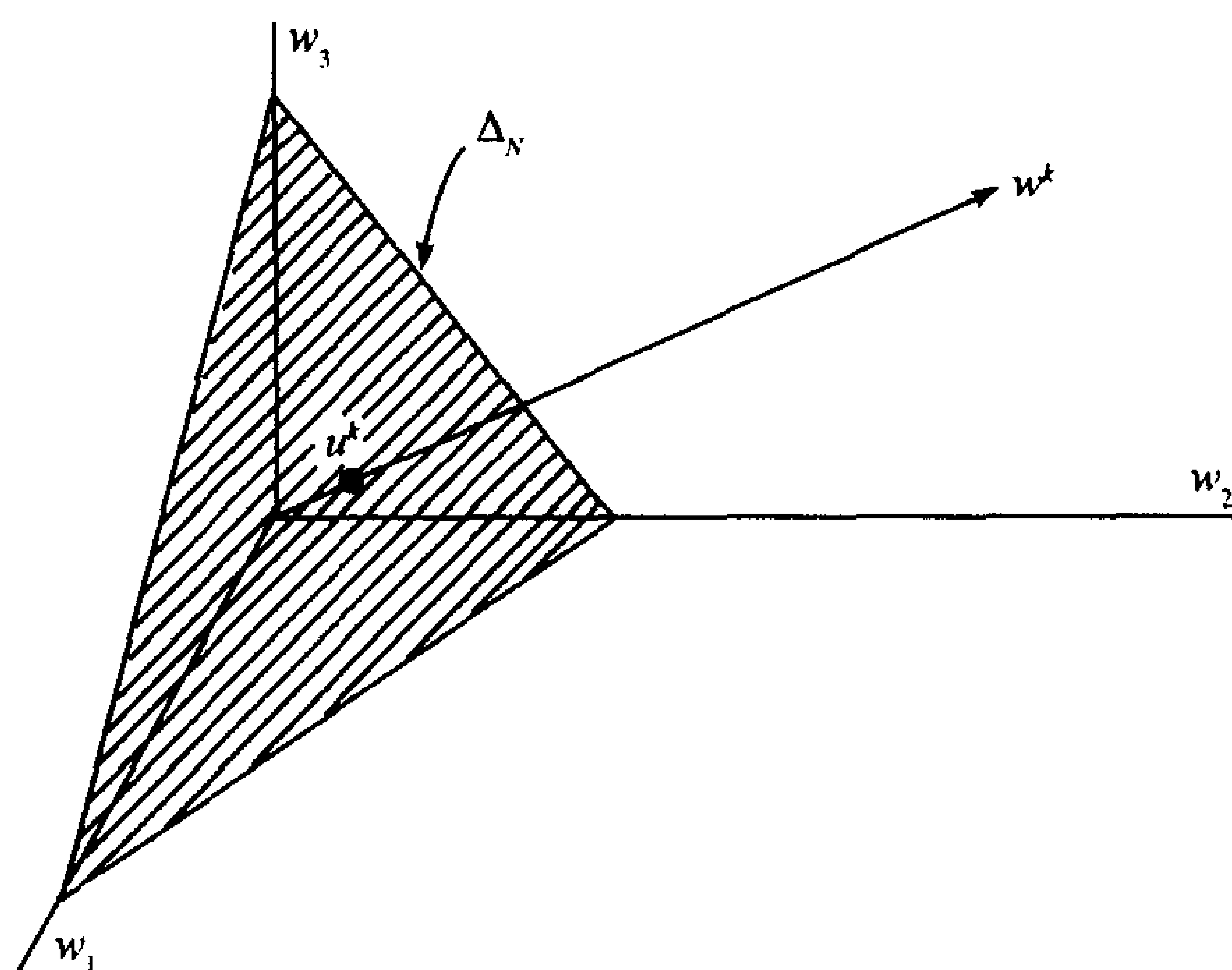


图 5-9 u^k 是 w^k 在单纯形 Δ_N 上的投影

定义 $T(0)$ 为 $\widetilde{\boldsymbol{w}}^k$ 的所有分量都同时发生所经过的世代数, 即使 $\widetilde{\boldsymbol{w}}^k$ 的所有分量都相等的最小 $k-1$. 这样一代称为**等分量代**. 213

引理 8.1 $\Pr\{T(0)<\infty\}=1$. 实际上, $E[T(0)]\leqslant N^N$.

证明 令 $\boldsymbol{w}^1=(w_1^1,\cdots,w_N^1)$ 是第一代的适合度. 使 $T(0)=1$ 出现的一种方式是要后代群体有

$$\widetilde{w}_k^2=w_v^1, \quad 对所有 k,$$

其中

$$w_v^1=\max\{w_1^1,\cdots,w_N^1\}.$$

显然, 这个事件发生的概率至少是 $\alpha=(1/N)^N$. 于是 $\Pr\{T(0)=1\}\geqslant\alpha$, 且 $\Pr\{T(0)>1\}\leqslant 1-\alpha$.

对从第二代到第三代的变迁使用同样的估计. 故

$$\Pr\{T(0)>2|T(0)>1\}\leqslant 1-\alpha,$$

且

$$\Pr\{T(0)>2\}\leqslant(1-\alpha)^2.$$

直接归纳, 导出 $\Pr\{T(0)>k\}\leqslant(1-\alpha)^k$, 于是 $E[T(0)]\leqslant 1/\alpha=N^N$. ■

令 $T(0)+T(1)$ 是 $T(0)$ 以后第一次出现等分量代的世代序数, 定义 $T(0)+\cdots+T(k)$ 是 $T(0)+\cdots+T(k-1)$ 以后第一次出现等分量代的世代序数. 关键的是, 过程 $\{u^k\}$ 在第 $T(0)$、$T(0)+T(1)$、$\cdots$ 世代重新开始, 因而这一正整数值随机变量序列形成一个更新过程. 为说明其中原由, 设 $\widetilde{w}$ 是第 $T(k)$ 代的公共适合度. 那么, $\widetilde{w}^{T(k)+1}$ 214

的所有分量具有值 $\widetilde{w}$, $\boldsymbol{w}^{T(k)+1}$ 的第 j 个分量是 $\widetilde{w}V_j^{T(k)+1}$. $\widetilde{w}$ 的影响在下一步消失, 依照 (8.1), 因为多项分布概率

$$u_j = \frac{\widetilde{w}V_j^{T(k)+1}}{\sum_{i=1}^{N}\widetilde{w}V_i^{T(k)+1}} = \frac{V_j^{T(k)+1}}{\sum_{i=1}^{N}V_i^{T(k)+1}}, \quad j = 1, 2, \cdots, N,$$

与 $\widetilde{w}$ 无关. 由此, 对于随后的每一等分量世代, 过程唯一地取决于独立同分布随机变量组的变异系数. 由此推知, 序列 $T(0), T(1), \cdots$ 成为一个延迟 (滞后) 更新过程.

设 $N(t)$ 表示由 $\{T(k)\}$ 导出的更新计数过程, $S_k = T(0) + \cdots + T(k-1)$, 其中 $S_0 = 0$. 选取世代序数 m, 则 $S_{N(m)}$ 是 m 世代以前的所有变异之前的适合度都相同的最后一个世代. 相对适合度 $\boldsymbol{u}^m = (u_1^m, \cdots, u_n^m)$ 的分布是什么?

在时刻 $S_{N(m)}$, 相对适合度有相等的分量 $1/N$. 考虑

$$\delta_m = m - S_{N(m)},$$

它是 $S_{N(m)}$ 与 m 之间所经过的世代数. 变异作用的多项分布规律在一整个世代内进行选择, 它可以用转移分布函数 Γ 来概括:

$$\Gamma(z_1, \cdots, z_N; \eta_1, \cdots \eta_N) = \Pr\{(\eta_1, \cdots, \eta_N) \text{ 的相对适合度变成 } (\eta_1', \cdots, \eta_N'), \text{ 其中 } \eta_k' \leqslant z_k, k = 1, \cdots, N\}.$$

因此,

$$\Pr\{u_k^m \leqslant z_k, k = 1, \cdots, N | \delta_m = 1\} = \Gamma\left(z_1, \cdots, z_N; \frac{1}{N}, \cdots, \frac{1}{N}\right).$$

从某一等分量世代开始, 在变异和抽样选择作用下, k 代繁殖过程的结果可用如下公式表示:

$$\Pr\{u_j^m \leqslant z_j, j = 1, \cdots, N | \delta_m = k\} = \Gamma^{(k)}\left(z_1, \cdots, z_N; \frac{1}{N}, \cdots, \frac{1}{N}\right),$$

215

其中, $\Gamma^{(k)}$ 表示由变换 Γ 导出的 k 重复合转移分布. 利用全概率公式得

$$\Pr\{u_j^m \leqslant z_j, j = 1, \cdots, N\} = \sum_{k=1}^{m} \Pr\{\delta_m = k\} \Gamma^{(k)}\left(z_1, \cdots, z_N; \frac{1}{N}, \cdots, \frac{1}{N}\right) + \Pr\{T(0) \geqslant m\} \Gamma^{(m+1)}(z_1, \cdots, z_N; u_1^1, \cdots, u_N^1). \tag{8.3}$$

于是, 由更新过程 $\{T(k)\}_{k=1}^{\infty}$ 的现龄极限分布推出

$$\lim_{n\to\infty} \Pr\{\delta_n = k\} = \frac{1 - \Pr\{T(1) \leqslant k\}}{E[T(1)]}, \quad ((6.5) \text{ 的特殊情况}).$$

结合 (8.2), 导出

$$\lim_{m\to\infty}\Pr\{u_j^m\leqslant z_j, j=1,\cdots,n\}$$
$$=\sum_{k=1}^{\infty}\left\{\frac{1-\Pr\{T(1)\leqslant k\}}{E[T(1)]}\right\}\Gamma^{(k)}\left(z_1,\cdots,z_N;\frac{1}{N},\cdots,\frac{1}{N}\right).$$

(读者容易验证, 极限与求和号可交换.)

最后的关系式描述了在群体进化时相对适合度的极限性质, 实际上提供了平稳分布的一个显式公式.

B. 分支过程

设在时刻 $t=0$ 有单个生物, 它的寿命是服从分布 F 的随机时间 T_0. 它的一生以概率 p_j 生出 j 个新的生物, $j=1,2,\cdots$. 每个新的生物的生活和繁殖都独立于群体的其他成员, 并且它们的寿命具有同一分布. 设 $m=\sum jp_j$ 为后代分布的均值. 令 $M(t)$ 表示时刻 t 活着的生物体个数的均值. 在第一代生物死亡之后, 应用全概率法则及更新论证法 (详细内容可参见第 8 章), 我们得

$$M(t)=1-F(t)+\sum_{j=1}^{\infty}p_j\int_0^t jM(t-x)\mathrm{d}F(x)$$
$$=1-F(t)+m\int_0^t M(t-x)\mathrm{d}F(x). \tag{8.4}$$

216

若无因子 m, 方程 (8.4) 即为一个更新方程. 当 $m>1$ 时, 我们可以将 (8.4) 变换成正常的更新方程. 取 β, 使其满足 $\int_0^{\infty}\mathrm{e}^{-\beta x}\mathrm{d}F(x)=1/m$. 由于 $\int_0^{\infty}\mathrm{e}^{-\lambda x}\mathrm{d}F(x)$ 是 λ 的严格递减的连续函数, 且当 $\lambda=0$ 时取值为 1, 当 $\lambda\to\infty$ 时趋于 0, 所以上述的 $\beta>0$ 是唯一存在的.

定义

$$\hat{F}(t)=m\int_0^t\mathrm{e}^{-\beta x}\mathrm{d}F(x),$$
$$\hat{M}(t)=\mathrm{e}^{-\beta t}M(t),$$

以及

$$g(t)=\mathrm{e}^{-\beta t}[1-F(t)].$$

方程 (8.4) 乘以 $\mathrm{e}^{-\beta t}$, 得

$$\mathrm{e}^{-\beta t}M(t)=\mathrm{e}^{-\beta t}[1-F(t)]+\int_0^t\mathrm{e}^{-\beta(t-x)}M(t-x)m\,\mathrm{e}^{-\beta x}\mathrm{d}F(x),$$

用新的记号改写成

$$\hat{M}(t) = g(t) + \int_0^t \hat{M}(t-x)\mathrm{d}\hat{F}(x).$$

直接验证可知 $g(t)$ 是直接黎曼可积的, 于是, 若 F 不是算术的, 由更新定理

$$\lim_{t\to\infty} \hat{M}(t) = \frac{\displaystyle\int_0^\infty g(x)\mathrm{d}x}{\displaystyle\int_0^\infty x\mathrm{d}\hat{F}(x)}.$$

此时, $\displaystyle\int_0^\infty g(t)\mathrm{d}t = (m-1)/\beta m$, 推得

$$\lim_{t\to\infty} \mathrm{e}^{-\beta t} M(t) = \frac{m-1}{\beta m^2 \displaystyle\int_0^\infty x\mathrm{e}^{-\beta x}\mathrm{d}F(x)},$$

因此当 t 充分大时, $M(t)$ 按比率 $\mathrm{e}^{\beta t}$ 指数增加. 参数 β 就是所谓 Malthusian 群体增长率.

217

C. 存储理论

某老板手头必须保持一定的库存商品. 当存货不足时, 他定购新商品以补充他的库存. 实施中的库存策略假定是 (s, S) 型的. (这是一种普通的作法). 明确地说, 两个水平 $s < S$ 是事先规定的. 假设最初商品在水平 S, 又明确规定一个周期的长度, 并要求在每个周期之末盘点存货. 若在某个周期结束时, 货存水平低于 s, 则将发出订货单以使库存水平上升到 S 以备下一周期开始供应市场.

设 X_i 是第 i 个周期的累计需求量. 我们假定 $X_1, X_2, \cdots$ 是独立同分布的正随机变量, 分布函数为 F. 设 $N(t)$ 是相应的更新计数过程. 显然, $N(S-s)+1$ 是首次为补充库存而发出订单之前的需求周期数, 发出订单以后商品库存又恢复到水平 S. 略作思考即可发现, 我们正涉及两个过程: 第一个是需求过程; 第二个是补充库存过程.

设在第 $i-1$ 次和第 i 次补充库存之间的需求周期数是 θ_i, 则 $\{\theta_i\}$ 是离散 (整数值) 的更新过程, 其均值为 $E(\theta_i) = 1 + M(S-s)$, 且

$$\Pr\{\theta_i = k\} = F_{k-1}(S-s) - F_k(S-s). \tag{8.5}$$

令 W_n 是第 n 个需求周期末的存货水平. 定义 G_n 为条件分布

$$G_n(x) = \Pr\{S - x \leqslant W_n | s \leqslant W_n\}.$$

这是在已知第 n 个周期末存货水平不低于 s 的条件下第 n 个周期末的存货水平的分布. 这个分布可以在设定 δ_n 下计算, δ_n 是最后一次补充库存至 S 之前的需求周期数. (回想一下, 我们假设在时刻 0, 商品货存量是 S.) 我们得到

$$\begin{aligned}G_n(x)&=\sum_{j=1}^{\infty}\Pr\{\delta_n=j\}\Pr\{X_1+\cdots+X_j\leqslant x|X_1+\cdots+X_j\leqslant S-s\}\\&=\sum_{j=1}^{\infty}\Pr\{\delta_n=j\}\Big(\frac{F_j(x)}{F_j(S-s)}\Big).\end{aligned}\tag{8.6}$$

但从 (8.5) 知,

$$\Pr\{\theta_1\leqslant j\}=1-F_j(S-s).\tag{8.7}$$

218

因此, 借助于更新过程 $\{\theta_i\}$ 的剩余寿命随机变量的极限定理 (见 (6.2)) 及式 (8.7), 我们推得

$$\begin{aligned}\lim_{n\to\infty}\Pr\{\delta_n=j\}&=\frac{1}{E[\theta_1]}\Pr\{\theta_1>j\}=\frac{F_j(S-s)}{E[\theta_1]}\\&=\frac{F_j(S-s)}{1+M(S-s)}.\end{aligned}\tag{8.8}$$

结合 (8.6) 和 (8.8), 我们推得

$$\begin{aligned}\lim_{n\to\infty}\Pr\{S-x\leqslant W_n|s\leqslant W_n\}&=\lim_{n\to\infty}\sum_{j=1}^{\infty}\Pr\{\delta_n=j\}\frac{F_j(x)}{F_j(S-s)}\\&=\frac{M(x)}{1+M(S-s)},\end{aligned}$$

这给出了不急于进货周期存货水平的极限分布.

D. 泊松过程的特征

泊松过程是一相当特别的更新过程. 本节将提出泊松过程作为更新过程族中的特殊过程的若干特性. 为此, 我们将需要利用 5.6 节中的几个极限定理.

设 $\{X_k\}$ 是一更新过程, $E[X_k]=\mu<\infty$, 且 $F(x)=\Pr\{X_k\leqslant x\}$. 假定 $F(0)=0$. 定义

$$F_t(x)=\begin{cases}F(x), & 对\ 0\leqslant x<t,\\ 1, & 对\ t\leqslant x.\end{cases}$$

[与 (3.5) 对照]. 显然, $F_t(x)$ 是 $\min\{X_k,t\}$ 的分布函数.

定理 8.1 (a) 若存在序列 $\{t_j\}$, 当 $j\to\infty$ 时 $t_j\to\infty$, 且现龄 δ_t 满足

$$F_{t_j}(x)=\Pr\{\delta_{t_j}\leqslant x\},\quad 对所有\ x,$$

则 F 是指数分布.

(b) 若存在序列 $\{t_j\}$, 当 $j \to \infty$ 时 $t_j \to \infty$, 且

$$F(x) = \Pr\{\gamma_{t_j} \leqslant x\}, \quad \text{对所有 } x,$$

219 [与 (3.3) 对照], 则 F 是指数分布.

证明 由于 (a) 与 (b) 十分类似, 故我们只证明 (a). 据 (6.5) 的结论, 现龄 δ_t 的极限分布应是

$$\lim_{t\to\infty} \Pr\{\delta_t > y\} = \mu^{-1} \int_y^\infty \{1 - F(z)\}\mathrm{d}z.$$

让 t 按 t_j 增长 (见定理的假设前提), 我们导出函数方程

$$1 - F(y) = \mu^{-1} \int_y^\infty \{1 - F(z)\}\mathrm{d}z.$$

右端式子显然对 y 是可微的, 因而得到初等的一阶微分方程

$$\frac{\mathrm{d}}{\mathrm{d}y}\{1 - F(y)\} = -\frac{1}{\mu}\{1 - F(y)\},$$

满足 $F(0) = 0$ 的解是

$$1 - F(y) = \mathrm{e}^{-\lambda y}, \quad \lambda = 1/\mu,$$

即为所证. ■

定理 8.2 设对某 $t_0 > 0$, 成立

$$\Pr\{\delta_t \leqslant x\} = F_t(x), \quad 0 \leqslant t < t_0, x \geqslant 0.$$

则对某 $\lambda > 0$, 使得 $F(x) = 1 - \mathrm{e}^{-\lambda x}, 0 \leqslant x < t_0$.

证明 对 $0 \leqslant x < t$, 有

$$\begin{aligned}
\Pr\{\delta_t \leqslant x\} &= \sum_{j=1}^\infty \Pr\{\delta_t \leqslant x \text{ 且 } N(t) = j\} \\
&= \sum_{j=1}^\infty \Pr\{t - x < S_j \leqslant t \text{ 且 } S_{j+1} > t\} \\
&= \sum_{j=1}^\infty \int_{t-x}^t [1 - F(t-y)]\mathrm{d}F_j(y) \\
&= \int_{t-x}^t [1 - F(t-y)]\mathrm{d}M(y).
\end{aligned}$$

于是, 由假设知, 对 $0 \leqslant x \leqslant t < t_0$, 成立

$$F(x) = \int_{t-x}^{t} [1 - F(t-y)]\mathrm{d}M(y), \tag{8.9}$$

且

$$\frac{1}{x}F(x) = \frac{1}{x}\int_{t-x}^{t} [1 - F(t-y)]\mathrm{d}M(y). \tag{8.10}$$

220

函数 $M(t)$ 是有限、右连续、不减的, 并且对稠于任一区间 $(0, t_0]$ 的无限多个 t, 具有有限导数 $M'(t)$. 选择 $t = \tau$ 使 $M'(\tau) < \infty$. 则当 x 递减趋于 0 时, (8.10) 在 $t = \tau$ 有极限, 且

$$F'(0) = [1 - F(0)]M'(\tau) = M'(\tau).$$

但是, 因 (8.10) 的左端与 t 无关, 故对所有 t 右极限必须存在, 且对所有 $t < t_0$ 有

$$F'(0) = M'(t).$$

令 $\lambda = F'(0)$, 于是 $M(t) = \lambda t$. 将此代入 (8.9), 得

$$F(x) = \int_{t-x}^{t} [1 - F(t-y)]\lambda\mathrm{d}y.$$

对 x 微分, 得

$$\mathrm{d}F(x)/\mathrm{d}x = -\lambda[1 - F(x)], \quad 0 \leqslant x < t_0$$

或

$$F(x) = 1 - \mathrm{e}^{-\lambda x}, \quad 0 \leqslant x < t_0,$$

证毕. ■

5.9 更新过程的叠加

本节指出, 在一定条件下无限多一致稀疏更新过程的叠加是泊松过程. 正如中心极限定理表明正态分布在相当广泛条件下可用来描述某些随机变量那样, 定理 9.1 将给出在许多场合下泊松假设成立的重要依据.

我们还要特别介绍几个附属于更新过程叠加的其他结果, 以引导读者对当前某些热门的研究领域有所了解.

对每个整数 $n = 1, 2, \cdots$ 和每个 $i = 1, \cdots, k_n$, 其中当 $n \to \infty$ 时 $k_n \to \infty$, 令 $N_{ni}(t)$ 表示具有间隔时间分布为 $F_{ni}(t)$ 的更新计数过程. 集合 $\{N_{ni}(t); n = 1, 2, \cdots, i = 1, \cdots, k_n\}$ 构成随机过程的三角阵. 对每个 n, 我们假设过程 $\{N_{n1}(t)\}, \cdots, \{N_{nk_n}(t)\}$ 是独立的.

所谓叠加过程 $N_n(t)$ 是指一列计数过程之和:

221

$$N_n(t)=\sum_{i=1}^{k_n}N_{ni}(t),\quad t\geqslant 0.$$

但叠加过程一般不是更新计数过程, 这是由于其间隔时间不相互独立, 也不同分布. 事实上, "事件" 之间区间长度的分布往往是复杂而难以处理的.

定义 9.1 三角阵 $\{N_{ni}(t)\}$ 称为无限小的, 如果对于每个 $t\geqslant 0$ 有

$$\lim_{n\to\infty}\max_{1\leqslant i\leqslant k_n}F_{ni}(t)=0. \tag{9.1}$$

在证明主要定理之前, 先给出一个预备引理.

引理 9.1 设 $\{N_{ni}(t)\}$ 是间隔时间分布为 $\{F_{ni}(t)\}$ 的无限小阵. 令 $F_{ni}^{*j}(t)$ 表示 $F_{ni}(t)$ 的 j 重卷积. 若存在某个有限不减函数 $c(t)$, 对任意 t 满足

$$\limsup_{n\to\infty}\sum_{i=1}^{k_n}F_{ni}(t)\leqslant c(t), \tag{9.2}$$

则

(a) $\displaystyle\lim_{n\to\infty}\sum_{i=1}^{k_n}\sum_{j=2}^{\infty}[F_{ni}(t)]^j=0$,

(b) $\displaystyle\lim_{n\to\infty}\sum_{i=1}^{k_n}\sum_{j=2}^{\infty}F_{ni}^{*j}(t)=0$,

(c) $\displaystyle\lim_{n\to\infty}\sum_{i=1}^{k_n}F_{ni}^{*j}(t)=0$, 且关于 $j=2,3,\cdots$ 是一致的.

证明 我们仅证明 (a); (b) 和 (c) 易由不等式 $F_{ni}^{*j}(t)\leqslant[F_{ni}(t)]^j$[参见 (4.2)] 推得. 令

$$A_n(t)=\sum_{i=1}^{k_n}\sum_{j=2}^{k_n}[F_{ni}(t)]^j.$$

任给 $\varepsilon>0$ 和 $t>0$, 由于 $\{F_{ni}(t)\}$ 是无限小的, 故存在 n_0 使得对于 $i=1,\cdots,k_n$ 和任何 $n\geqslant n_0$ 有

$$F_{ni}(t)\leqslant\varepsilon.$$

于是, 对于 $n\geqslant n_0$ 且沿着 (9.2) 中取上极限的子列增加, 我们有

222

$$A_n(t) \leqslant \sum_{i=1}^{k_n} \sum_{j=2}^{\infty} \varepsilon^{j-1} F_{ni}(t)$$

$$= \frac{\varepsilon}{1-\varepsilon} \sum_{i=1}^{k_n} F_{n_i}(t)$$

$$\leqslant 2 \frac{\varepsilon c(t)}{1-\varepsilon}.$$

由于 $\varepsilon > 0$ 为任意的, 我们得到

$$\limsup_{n \to \infty} A_n(t) = 0.$$

■

定理 9.1 设 $\{N_{ni}(t)\}$ 是叠加过程 $N_n(t)$ 的更新过程的无限小阵. 则

$$\lim_{n \to \infty} \Pr\{N_n(t) = j\} = \frac{e^{-\lambda t}(\lambda t)^j}{j!}, \quad j = 0, 1, 2, \cdots, \tag{9.3}$$

当且仅当

$$\lim_{n \to \infty} \sum_{i=1}^{k_n} F_{ni}(t) = \lambda t. \tag{9.4}$$

证明 (1) 必要性 假设 (9.3) 成立. 对于 $j = 0$, 得

$$\lim_{n \to \infty} \Pr\{N_n(t) = 0\} = e^{-\lambda t},$$

或等价地,

$$\lim_{n \to \infty} [-\ln \Pr\{N_n(t) = 0\}] = \lambda t.$$

由条件知 $N_{n,1}(t), \cdots, N_{n,kn}(t)$ 是非负独立随机变量, 从而

$$-\ln \Pr\{N_n(t) = 0\} = -\ln \prod_{i=1}^{k_n} \Pr\{N_{ni}(t) = 0\}$$

$$= -\sum_{i=1}^{k_n} \ln[1 - F_{ni}(t)].$$

利用泰勒级数展开得

$$-\ln[1 - F_{ni}(t)] = \sum_{j=1}^{\infty} \frac{1}{j} [F_{ni}(t)]^j,$$

我们有

223

$$\lambda t=\lim_{n\to\infty}[-\ln\Pr\{N_n(t)=0\}]$$
$$=\lim_{n\to\infty}\left\{\sum_{i=1}^{k_n}F_{ni}(t)+\sum_{i=1}^{k_n}\sum_{j=2}^{\infty}\frac{1}{j}[F_{ni}(t)]^j\right\}.$$

由于上式第二项和式是非负的, 故

$$\limsup_{n\to\infty}\sum_{i=1}^{k_n}F_{ni}(t)\leqslant\lambda t,$$

再利用引理 9.1 推得第二项和式极限为 0, 故

$$\lambda t=\lim_{n\to\infty}\sum_{i=1}^{k_n}F_{ni}(t),$$

(9.4) 式证毕.

(2) 充分性 假设 (9.4) 成立. 我们对 m 使用归纳法证明

$$\lim_{n\to\infty}\Pr\{N_n(t)=m\}=\mathrm{e}^{-\lambda t}\frac{(\lambda t)^m}{m!},\quad m=0,1,\cdots.$$

步骤 1: $m=0$. 同上, 使用泰勒级数展开可得

$$\Pr\{N_n(t)=0\}=\prod_{i=1}^{k_n}[1-F_{ni}(t)]$$
$$=\exp\left[-\sum_{i=1}^{k_n}F_{ni}(t)-\sum_{i=1}^{k_n}\sum_{j=2}^{\infty}\frac{[F_{ni}(t)]^j}{j}\right].$$

利用假设 (9.4) 和引理 9.1, 上式指数的极限为 $-\lambda t$. 因此

$$\lim_{n\to\infty}\Pr\{N_n(t)=0\}=\exp\{-\lambda t\}.$$

步骤 2: 归纳推理步骤. 现假设命题对于 $m-1$ 成立, 即

$$\lim_{n\to\infty}\Pr\{N_n(t)=m-1\}=\frac{\mathrm{e}^{-\lambda t}(\lambda t)^{m-1}}{(m-1)!}.$$

以下证明命题对于 m 也成立. 令 $s_1,\cdots,s_r$ 是有限指标集合. 由无限小阵的定义知

224

$$\lim_{n\to\infty}\prod_{\substack{j=1\\ j\neq s_1,\cdots,s_r}}^{k_n}[1-F_{nj}(t)]=\mathrm{e}^{-\lambda t},\tag{9.5}$$

对每个固定 r, 此极限关于指标 $s_1, \cdots, s_r$ 是一致的. 现有

$$\Pr\{N_n(t)=m\} = P_n(t) + Q_n(t),$$

其中

$$P_n(t) = \Pr\{N_n(t)=m \text{ 和 } N_{ni}(t) \leqslant 1, i=1,\cdots,k_n\},$$

和

$$Q_n(t) = \Pr\{N_n(t)=m \text{ 和 } N_{ni}(t) \geqslant 2, \text{ 对于某个 } i\}.$$

现证明当 $n \to \infty$ 时 $Q_n(t) \to 0$. 事实上, 对任意 $\varepsilon > 0$, 若 n 足够大满足 $F_{ni} \leqslant \varepsilon$ 时, 我们有

$$\begin{aligned} Q_n(t) &\leqslant \sum_{i=1}^{k_n} \Pr\{N_{ni}(t) \geqslant 2\} \\ &\leqslant \sum_{i=1}^{k_n} [F_{ni}(t)]^2 \\ &\leqslant \varepsilon \sum_{i=1}^{k_n} F_{ni}(t), \end{aligned}$$

这样, 利用假设 (9.4), 我们有

$$\limsup_{n\to\infty} Q_n(t) \leqslant \varepsilon\lambda t,$$

既然 ε 是任意的, 故当 $n \to \infty$ 时, $Q_n(t) \to 0$. 下面来求 $P_n(t)$ 的极限值, 设 $I(m)$ 是所有满足 $1 \leqslant i_j \leqslant k_n$ 的可能组合 $(i_1, \cdots, i_m)$, 则

$$\begin{aligned} P_n(t) &= \Pr\{N_n(t)=m \text{ 和 } N_{ni}(t) \leqslant 1,\ i=1,\cdots,k_n\} \\ &= \sum_{I(m)} \Pr\{N_{ni_1}(t)=1,\cdots,N_{ni_m}(t)=1 \text{ 和 } N_{ni}(t)=0, i \notin I(m)\} \\ &= \frac{1}{m}\sum_{i=1}^{k_n} \Pr\{N_{ni}(t)=1, N_n(t)-N_{ni}(t)=m-1, \text{ 和 } N_{nj}(t) \leqslant 1, j=1,\cdots,k_n\} \\ &= \frac{1}{m}\sum_{i=1}^{k_n} \{F_{ni}(t) - F_{ni}^{*2}(t)\} R_{ni}(t) \end{aligned}$$

(此处利用了 $N_{ni}(t)$ 相互独立的假设) 其中

$$R_{ni}(t) = \Pr\{N_n(t)-N_{ni}(t)=m-1, N_{nj}(t) \leqslant 1, j=1,\cdots,k_n, j \neq i\}.$$

225

但由归纳假设及 (9.5) 知, $R_{ni}(t) \to e^{-\lambda t}(\lambda t)^{m-1}/(m-1)!$, 当 $n \to \infty$ 时, 此极限关于 i 是一致的. 因此,

$$\begin{aligned}\lim_{n\to\infty} P_n(t) &= \lim_{n\to\infty} \frac{1}{m}\sum_{i=1}^{k_n}\{F_{ni}(t) - F_{ni}^{*2}(t)\}R_{ni}(t)\\&= \lim_{n\to\infty} \frac{1}{m}\sum_{i=1}^{k_n} F_{ni}(t)e^{-\lambda t}(\lambda t)^{m-1}/(m-1)!\\&= e^{-\lambda t}(\lambda t)^m/m!,\end{aligned}$$

定理得证. ■

例 1 设 $F(t)$ 是满足 $F(0)=0$ 和 $F'(0)=\lambda>0$ 的分布函数. 令

$$F_{ni}(t) = F(t/n), \quad i=1,\cdots,n,$$

对所有 n, 设 $N_{ni}(t), i=1,\cdots,n$, 是间距分布为 F_{ni} 的相互独立的更新计数过程. 那么, $N_{ni}(t)$ 是三角阵. 而且由于

$$\lim_{n\to\infty}\max_{1\leqslant i\leqslant n} F_{ni}(t) = \lim_{n\to\infty} F(t/n) = 0,$$

推知此阵是无限小的. 为验证 (9.4), 我们计算

$$\begin{aligned}\lim_{n\to\infty}\sum_{i=1}^{n} F_{ni}(t) &= \lim_{n\to\infty} nF(t/n)\\&= t\lim_{n\to\infty}\frac{F(t/n)}{t/n}\\&= \lambda t.\end{aligned}$$

因此, 叠加过程 $N_n(t)$ 的分布收敛于泊松过程.

在本节结束时, 我们介绍作为更新过程组成部分的泊松过程的两个特征. 两个独立泊松过程之和仍然是泊松过程, 只不过后者的参数是前面的两个对应参数之和. 下面我们将指出, 在更新过程中只有泊松过程才具有上述性质.

定理 9.2 设 $N_1(t)$ 和 $N_2(t)$ 是两个独立更新过程, 且具有相同的间距分布 F(其均值记为 μ). 令 $N(t)=N_1(t)+N_2(t)$, 如果 $N(t)$ 也是更新过程, 则 $N_1(t), N_2(t)$ 和 $N(t)$ 都是泊松过程.

226

证明 设 H 是 $N(t)$ 的间距分布, 那么

$$\begin{aligned}1-H(x) &= \Pr\{N(x)=0\}\\&= \Pr\{N_1(x)=0, N_2(x)=0\}\\&= [1-F(x)]^2.\end{aligned}$$

令 $\gamma_1(t), \gamma_2(t)$ 和 $\gamma(t)$ 分别是过程 N_1, N_2 和 N 在时刻 t 的剩余寿命. 由于 N_1 和 N_2 组成 N, 因此必定有

$$\gamma(t) = \min\{\gamma_1(t), \gamma_2(t)\}$$

和

$$\Pr\{\gamma(t) > x\} = [\Pr\{\gamma_1(t) > x\}]^2$$

成立. 令 $t \to \infty$, 并利用式 (6.5) 中剩余寿命的渐近分布, 我们有

$$\frac{1}{\nu}\int_x^\infty [1 - H(y)]\mathrm{d}y = \frac{1}{\mu^2}\left\{\int_x^\infty [1 - F(y)]\mathrm{d}y\right\}^2, \tag{9.6}$$

其中 $\nu = \int_0^\infty [1 - H(y)]\mathrm{d}y$. 两边关于 x 均可微 [1], 需注意 $1 - H(x) = [1 - F(x)]^2$, 微分 (9.6) 得

$$\frac{1}{\nu}[1 - F(x)]^2 = \frac{1}{\nu}[1 - H(x)] = \frac{2}{\mu^2}\left\{\int_x^\infty [1 - F(y)]\mathrm{d}y\right\}[1 - F(x)],$$

或

$$1 - F(x) = \frac{2\nu}{\mu^2}\int_x^\infty [1 - F(y)]\mathrm{d}y.$$

记 $G(x) = 1 - F(x)$. 对上式求微分, 得

$$\frac{\mathrm{d}G(x)}{\mathrm{d}x} = -\frac{2\nu}{\mu^2}G(x),$$

在初始条件 $G(0) = 1$ 之下, 其解为

$$G(x) = 1 - F(x) = \mathrm{e}^{-\lambda x},$$

其中 $\lambda = 2\nu/\mu^2$. 这样, $N_1(t)$ 和 $N_2(t)$ 皆为泊松的, 从而, 其和 $N(t)$ 也是泊松的. ■

定理 9.3 设 $N_1(t)$ 是参数为 μ 的泊松过程, $N_2(t)$ 是具有有限平均间距的更新过程. 若 $N_1(t)$, $N_2(t)$ 相互独立且 $N(t) = N_1(t) + N_2(t)$ 是一更新过程, 则 $N_2(t)$必定也是泊松过程.

227

证明 我们使用定理 9.2 相同的技巧. 设 N_1, N_2 和 N 分别具有间距分布 $1 - \mathrm{e}^{-t/\mu}, G(t)$ 和 $H(t)$. 则

$$1 - H(t) = \{1 - G(t)\}\mathrm{e}^{-t/\mu}, \tag{9.7}$$

和

$$\frac{1}{\mu_H}\int_x^\infty [1 - H(y)]\mathrm{d}y = \frac{\mathrm{e}^{-x/\mu}}{\mu_G}\int_x^\infty [1 - G(y)]\mathrm{d}y,$$

1. 若 F 是连续的, 这是显然的. 一般情况需进一步论证.

其中 μ_H 和 μ_G 分别表示 H 和 G 的均值.

对上式求微分, 得

$$\frac{\mu_G}{\mu_H}[1-H(x)] = \mathrm{e}^{-x/\mu}[1-G(x)] + \frac{1}{\mu}\mathrm{e}^{-x/\mu}\int_x^\infty [1-G(y)]\mathrm{d}y,$$

结合 (9.7), 得

$$[1-G(x)]\Big[\frac{\mu_G}{\mu_H}-1\Big] = \frac{1}{\mu}\int_x^\infty [1-G(y)]\mathrm{d}y. \tag{9.8}$$

令

$$\lambda = \mu\Big[\frac{\mu_G}{\mu_H}-1\Big], \quad 和 \quad F(x) = 1-G(x).$$

对 (9.8) 求微分得

$$-\lambda\mathrm{d}F(x)/\mathrm{d}x = F(x),$$

或

$$F(x) = \mathrm{e}^{-x/\lambda},\ x \geqslant 0,$$

即得

$$G(x) = 1-\mathrm{e}^{-x/\lambda},\ x \geqslant 0,$$

定理证毕. ■.

初等问题

1. 若 $\Pr\{X_i=1\}=\dfrac{1}{3}, \Pr\{X_i=2\}=\dfrac{2}{3}$, 试计算

228

$$\Pr\{N(1)=k\}, \Pr\{N(2)=k\}, \Pr\{N(3)=k\}.$$

2. 一个病人到医院候诊室, 以概率 1/5 即刻就诊, 以概率 4/5 需延迟一个小时. 经过一个小时等候之后, 又以概率 1/5 即刻就诊, 以概率 4/5 再延迟一个小时, 如此等等.

(a) 首次到达等候时间的分布是什么?

(b) 假定到达病人数服从参数为 1 的泊松过程, 并且每次到达都遵循上述相同的模式, 试求至少需候诊 8 小时的病人数分布.

3. 设某地 A 的天气情况分晴天和雨天两种. 每次雨天持续的天数服从参数为 2 天的泊松分布, 而晴天是服从均值为 7 天的几何分布. 假定晴天和雨天持续的时间是彼此独立的. 试求长久之后, 在给定的一天中是雨天的概率为多少?

4. 设个体的随机寿命分布函数为 $F(x)$, 试求使用时间为 x 的个体平均剩余寿命.

解答:

$$\mathrm{e}(x) = E[X-x|X>x] = \frac{\displaystyle\int_x^\infty \{1-F(t)\}\mathrm{d}t}{1-F(x)}.$$

5. 设寿命分布函数 $F(x)$ 具有概率密度函数 $f(x)$. 函数 $r(x) = f(x)/[1-F(x)]$ 称为 hazard 比. 试证明使每单位时间长久平均费用

$$\theta(T) = \frac{c_1[1-F(T)] + c_2F(T)}{\int_0^T [1-F(x)]\mathrm{d}x}$$

达到最小的替换时间 T^* 必定满足

$$r(T^*) \times \int_0^{T^*} [1-F(x)]\mathrm{d}x - F(T^*) = \frac{c_1}{c_2 - c_1}.$$

6. 假设汽车相继到达一大门前, 每辆汽车的随机长度 L 具有分布函数 $F(\xi)$. 第一辆汽车到达并在大门前停车. 以后每辆汽车相继到达并以一定距离停靠在前一辆汽车的后面, 此距离服从 [0,1] 上均匀分布. 考虑这一串排成队的汽车, 记 N_x 为距离大门不超过 x 的汽车数. 若 (i) $F(\xi)$ 为退化分布, 即 $F(\xi) = c$; (ii) $F(\xi) = 1 - \mathrm{e}^{-\xi}$, 试分别确定其

$$\lim_{x\to\infty} E[N_x]/x.$$

229

解答：$2/(1+2c)$ 和 $2/3$.

7. 设在每天初始的时间段内顾客到达出租汽车站的时刻排成分布律为 $F(x)$ 的更新过程. 假定出租汽车有无限多, 且每个顾客在站上所付的钱服从分布律 $G(x)$, $x > 0$.

(i) 试写出在一天的时刻 t 之前出租汽车站收入总和的表达式.

(ii) 试确定极限

$$\lim_{t\to\infty} E[\text{在第一个 } t \text{ 时间内的收入}]/t.$$

8. 考虑型 II 的计数器. 每次脉冲到达之后闭锁固定的 τ 单位时间. 假定脉冲到达遵循参数为 λ 的泊松分布. 试确定计数器在时刻 t 开放的概率 $p(t)$.

解答：

$$p(t) = \begin{cases} \mathrm{e}^{-\lambda t}, & t < \tau, \\ \mathrm{e}^{-\lambda\tau}, & t \geqslant \tau. \end{cases}$$

问　　题

1. 设一更新过程具有寿命密度

$$f(x) = \begin{cases} p\mathrm{e}^{-p(x-\delta)}, & x > \delta, \\ 0, & x \leqslant \delta, \end{cases}$$

其中 $\delta > 0$ 是固定的. 试求 $\Pr\{N(t) \geqslant k\}$.

2. 设某生物体寿命是一具有分布函数 $F(x)$ 的随机变量, 在其一生中依照强度函数为 $\lambda(u)$ 的非齐次泊松过程繁殖后代. 每一后代又独立地依照同样概率规律产生其后代. 这样, 群体不断衍生发展. 若

$$1 < \int_0^\infty \{1-F(u)\}\lambda(u)\mathrm{d}u < \infty,$$

试证明平均群体总量 $m(t)$ 渐近按速率为 $r>0$ 的指数增长, 其中 r 是如下方程的唯一解:

230

$$1=\int_0^{\infty} \mathrm{e}^{-ru}\{1-F(u)\}\lambda(u)\mathrm{d}u.$$

提示: 设 $B(t)$ 是至时刻 t 为止所繁殖个体的平均数, 先导出关于 $B(t)$ 的更新方程, 由此推出 $B(t)$ 依速率 r 指数增长, 然后用 $B(u)$ 表达 $m(t)$, 其中 $u \leqslant t$.

3. 试证明 $\lim\limits_{t\to\infty} V(t)/t=\sigma^2/\mu^3$, 其中 $V(t)$ 是更新过程 $N(t)$ 的方差, μ 和 $\sigma^2<\infty$ 分别是间距分布的均值和方差.

4. 对于具有分布 $F(x)$ 的更新过程, 试计算

$$p(t)=\Pr\{\text{在 } (0,t] \text{ 更新次数为奇数}\}.$$

(i) 当 $F(x)$ 是参数为 λ 的泊松分布时; (ii) 当 $F(t)=\int_0^t x\mathrm{e}^{-x}\mathrm{d}x$ 时, 分别具体计算之.

5. 沿着一对染色体的纵向依照参数为 λ 的泊松过程出现破裂和重新组合. 为说明清楚, 设破裂出现在点 t_1 和 t_2



则重新组合的染色体具有形式



试确定距离为 l 的两点 A, B, 当重新组合后仍然在相同染色体上结合的概率.

6. 试证明对应于寿命密度为

$$f(x)=\lambda^2 x\mathrm{e}^{-\lambda x}, \quad x \geqslant 0$$

的更新函数是

$$M(t)=\frac{1}{2}\lambda t-\frac{1}{4}(1-\mathrm{e}^{-2\lambda t}).$$

7. 在整体替换模型中, 设 c_1 是计划替换的费用, c_2 是因元件损坏而替换所需的费用. 利用长久运行每单位时间平均费用公式 $[c_1+c_2M(T)]/T$, 证明使花费最小的整体替换时间 T^* 应满足

$$\mathrm{e}^{-2\lambda T^*}(1+2\lambda T^*)=1-(4c_1/c_2),$$

其中 $c_2>4c_1$, 并且其寿命密度恰为问题 6 中的 $f(x)$.

8. 设 $X_1, X_2, \cdots,$ 是独立同分布随机变量, 且服从在 (0,1) 上均匀分布. 令 N_k 是满足 $X_1^k+X_2^k+\cdots+X_n^k \leqslant 1<X_1^k+\cdots+X_{n+1}^k$ 的最小下标 n. 试求

231

$$\lim_{k\to\infty} E(N_k)/k.$$

提示：证明并利用恒等式 $E(S_{N_k+1}) = E(N_k+1)E(X_1^k)$, 其中 $S_r = X_1^k + \cdots + X_r^k$, $r = 1, 2, \cdots$.

9. 试确定泊松过程全寿命 β_t 的分布.

答案：$\Pr\{\beta_t \leqslant x\} = 1 - [1 + \lambda \min\{t, x\}]\mathrm{e}^{-\lambda x}$.

10. 试证明在更新过程中, 被视为一个随机过程的现龄 $\{\delta_t; t \geqslant 0\}$ 是一马氏过程, 并求其转移分布函数

$$F(y; t, x) = \Pr\{\delta_{s+t} \leqslant y | \delta_s = x\}.$$

11. 设 $A(t)$ 是更新方程 $A(t) = a(t) + \int_0^t A(t-y)\mathrm{d}F(y)$ 的解, 其中 $a(t)$ 是满足 $a(0) = 0$ 的有界不减函数. 试证明 $\lim\limits_{t\to\infty} A(t)/t = a^*/\mu$, 其中 $a^* = \lim\limits_{t\to\infty} a(t)$ 且 $\mu < \infty$ 是 $F(x)$ 的均值.

12. 设一系统有两个状态："开" 与 "关". 任一时刻系统仅处于其中一个状态. 在时刻 0 它处于 "开" 状态, 在关闭之前它运行的随机时间 $\mathrm{T}_{开}$ 具有分布函数 $1 - \mathrm{e}^{-t\lambda}$. 继而进入 "关" 状态, 其间随机时间 $\mathrm{T}_{关}$ 具有相同分布函数 $1 - \mathrm{e}^{-t\lambda}$. 之后系统进行修理. 处于上述 "开"、"关" 两种状态的随机时间是彼此独立且同分布. 如此循环交替. 令 $W(t)$ 表示在时间区间 $(0, t)$ 内系统运行 (即处于 "开" 状态) 的全部时间, 试确定 $W(t)$ 的均值.

13. 对具有密度函数为

$$f(x) = \begin{cases} x\mathrm{e}^{-x}, & x \geqslant 0, \\ 0, & x \leqslant 0, \end{cases}$$

的随机变量进行连续独立的观察, 直到观察次数之和超过 t. 其中 $N+1$ 为所需的观察次数. 试证明

$$\Pr\{N = n\} = \frac{t^{2n+1}\mathrm{e}^{-t}}{\Gamma(2n+2)} + \frac{t^{2n}\mathrm{e}^{-t}}{\Gamma(2n+1)}.$$

14. 更新过程是记录在 $(0, t]$ 出现的点的数目的非负整数值随机过程, 而点出现的间隔时间是独立同分布, 其共同分布函数为 $F(x)$(当 $x \geqslant 0$) 或为 0(当 $x < 0$), 且 $F(x)$ 在 $x = 0$ 是连续的. 所谓修改更新过程是指上述间隔时间分布 $F(x)$ 在 $x = 0$ 具有跳跃 q 的一种更新过程. 试证明修改更新过程等价于一个普通更新过程, 但其中在每次到达的所记录点的数目是独立同分布的随机变量 $R_1, R_2, \cdots$; 其分布为

$$\Pr\{R_i = n\} = pq^{n-1}, \quad n = 1, 2, \cdots,$$

此处 $i = 1, 2, \cdots$, $p = 1 - q$.

232

15. 考虑具有间距分布函数为 $F(x)$ 的更新过程. 令 W 表示这样一个时间, 使得从上一次更新事件发生至此时的时间长度第一次超过 $\xi > 0$(固定常数). 试确定

$$V(t) = \Pr\{W \leqslant t\}$$

所满足的积分方程并计算 $E[W]$. (假设有一个事件发生在 $t = 0$)

16. 考虑间距分布为 $F(x)$ 的更新过程 $N(t)$, 记 $m_k(t)=E[N(t)^k]$. 证明 $m_k(t)$ 满足更新方程

$$m_k(t)=z_k(t)+\int_0^t m_k(t-\tau)\mathrm{d}F(\tau),\quad k=1,2,\cdots,$$

其中

$$z_k(t)=\int_0^t\sum_{j=0}^{k-1}\binom{k}{j}m_j(t-\tau)\mathrm{d}F(\tau).$$

提示：利用更新论证法.

17. (续上) 用归纳法证明

$$z_k(t)=(-1)^{k-1}\left[F(t)-\binom{k}{1}m_1(t)+\cdots+(-1)^k\binom{k}{k-1}m_{k-1}(t)\right].$$

18. 考虑仅有两个状态 A 和 B 的随机过程 $X(t), t\geqslant 0$, 用 $\xi_1,\eta_1,\xi_2,\eta_2,\cdots$ 分别表示状态相继逗留在 A 和 B 的时间, 并设 $X(0)$ 处于状态 A. 假定 $\xi_1,\xi_2,\cdots$ 是独立同分布随机变量, 其分布为 $F(\xi),\eta_1,\eta_2,\cdots$ 也是独立同分布随机变量, 其分布为 $G(\eta)$. 用 $Z(t)$ 和 $W(t)$ 分别表示在时间区间 $(0,t)$ 内处于状态 A 和 B 的总逗留时间. 显然, $Z(t)$ 和 $W(t)$ 是随机变量且 $Z(t)+W(t)=t$. 令 $N(t)$ 是由 $\xi_1,\xi_2,\cdots$ 产生的更新过程. 定义

$$\theta(t)=\eta_1+\eta_2+\cdots+\eta_{N(t)}.$$

试证明

$$\mathrm{P}\{W(t)\leqslant x\}=\mathrm{P}\{\theta(t-x)\leqslant x\},$$

并使用分布 F 和 G 表达此式.

答案：

$$\Pr\{W(t)\leqslant x\}=\sum_{n=0}^{\infty}G_n(x)[F_n(t-x)-F_{n+1}(t-x)],$$

233 其中 G_n 和 F_n 是通常意义下的卷积.

19. 考虑间距分布为 $F(x)$ 的更新过程, 假设每个事件以概率 $1-q$ 消失, 并以因子 $1/q$ 扩大时间尺度. 试证明所形成事件列构成一个更新过程, 其间距分布为

$$\sum_{n=1}^{\infty}(1-q)^{n-1}qF_n(x/q)=F(x;q),$$

其中 F_n 仍表示 F 的 n 重卷积.

20. (续上) 在前面问题中, 设 $\phi(s)$ 是 $F(x)$ 的拉普拉斯变换. 试求 $F(x;q)$ 的拉普拉斯变换.

答案：

$$\phi(s;q)=\frac{q\phi(sq)}{1-(1-q)\phi(sq)}.$$

21. (续上) 若 F 具有二阶矩, 试证明当 $q \to 0+$ 时, 对所有 s, $\mathrm{Re}\, s \geqslant 0$, 有

$$\phi(s;q) \to \frac{\lambda}{\lambda+s},$$

其中 $\lambda^{-1} = \int_0^\infty x \mathrm{d}F(x)$.

22. (续上) 利用第 1 章收敛性定理证明

$$F(x;q) \to 1 - \mathrm{e}^{-\lambda x}, \quad 当\ q \to 0+.$$

23. 考虑间距分布为 $G_0(x)$ 的更新过程. 假定每个事件以概率 q 保留, 以概率 $1-q$ 除掉, 并且时间尺度以因子 $1/q$ 扩大 (见问题 19). 试证明对于原来的和新的过程, 其平均间隔时间是相同的. 重复上述去除事件和扩大时间尺度的步骤, 从而得到一列更新过程, 其中原过程经过这样 n 次变换之后所得更新过程的间距分布记为 $G_{(n)}(x)$. 在所有这些变换中 q 是固定的. 试证明若 $0<q<1$, 则

$$\lim_{n\to\infty} G_{(n)}(x) = 1 - \mathrm{e}^{-x\mu},$$

其中 $\mu = \int_0^\infty \{1 - G_{(0)}(\xi)\}\mathrm{d}\xi$.

提示: 置

$$\phi_0(s) = \int_0^\infty \mathrm{e}^{-s\xi}\mathrm{d}G_0(\xi), \quad \phi_i(s) = \int_0^\infty \mathrm{e}^{-s\xi}\mathrm{d}G_i(\xi).$$

用归纳法证明

$$\phi_n(s) = \frac{q^n\phi_0(sq^n)}{1-(1-q^n)\phi_0(sq^n)}.$$

234

令 $n\to\infty$ 即导出与问题 20~22 相同的结果.

24. 设有同分布的更新过程三角阵 $N_{ni}(t)$, $l \leqslant i \leqslant n$, 其中间隔时间具有均值为 μ 的分布 $F(t)$. 考虑该阵的第 n 排, 在此排的每个过程中, 均以概率 $1/n$ 保留一个事件, 以概率 $1-(1/n)$ 抛弃一个事件, 这个程序独立地应用于所有事件. 通过这样删除手续所得到的一个新的更新过程三角阵, 记为 $N_{ni}^*(t)$. 然后把它们叠加得

$$N_n^*(t) = \sum_{j=1}^n N_{nj}^*(t), \quad 1 \leqslant n < \infty.$$

试证明

$$\lim_{n\to\infty} \Pr[N_n^*(t) = j] = \frac{\mathrm{e}^{-t/\mu}}{j!}(t/\mu)^j,$$

当且仅当 $F(t) = 1 - \mathrm{e}^{-t/\mu}$. 换言之, 叠加过程收敛于泊松过程当且仅当原有的更新过程是泊松的.

答案: 首先证明新阵 $N_{nj}^*(j)$ 是无限小的, 即证明

$$\lim_{n\to\infty}\left[\sup_{1\leqslant j\leqslant n} F_{nj}^*(t)\right] = 0,$$

其中 $F_{ni}^*(t)$ 是转换过程 $N_{ni}^*(t)$ 的间距分布. 事实上, 按问题 19 的论证方法, 我们有

$$\begin{aligned} F_{ni}^*(t) &= \sum_{j=1}^{\infty}\left(1-\frac{1}{n}\right)^{j-1}\frac{1}{n}F_j(t) \\ &\leqslant \frac{1}{n}\sum_{j=1}^{\infty}F_j(t) \leqslant \frac{1}{n}\sum_{j=1}^{\infty}[F(t)]^j \\ &\leqslant \frac{1}{n}\frac{F(t)}{1-F(t)}. \end{aligned}$$

其中若 $F(t)<1$, 显然有 $\lim_{n\to\infty}\left[\sup_{1\leqslant i\leqslant n}F_{ni}^*(t)\right]=0$. 若 $F(t)=1$, 我们可以确定一个适当的 j 使得 $F_j(t)<1$, 从而可得到类似结果.

进一步应用叠加定理 9.1, 得

$$\begin{aligned} \lim_{n\to\infty}\sum_{i=1}^{n}F_{ni}^*(t) &= \lim_{n\to\infty}\sum_{j=1}^{\infty}\left(1-\frac{1}{n}\right)^{j-1}F_j(t) \\ &= \sum_{j=1}^{\infty}F_j(t) \\ &= M(t) \quad \text{(更新函数)} \\ &= t/\mu, \end{aligned}$$

[235] 这等价于 $F(t)=1-\mathrm{e}^{-t/\mu}$.

25. 已知一具有有限均值的更新过程, 假设对所有 t 剩余寿命 γ_t 和现龄 δ_t 是独立随机变量, 试证明此过程是泊松的.

提示: 在恒等式

$$\Pr\{\delta_t>x,\gamma_t>y\}=\Pr\{\delta_t>x\}\Pr\{\gamma_t>y\},$$

中, 利用 5.5 节中的极限定理, 可导出关于

$$v(x)=\frac{1}{\mu}\int_x^{\infty}[1-F(\xi)]\mathrm{d}\xi,$$

的函数方程, 由此推得 $1-F(\xi)=\mathrm{e}^{-\xi/\mu}$.

26. (a) 若一批订货单依照参数为 λ 的泊松过程到达某中心办事处, 假设完成每一订货单任务需要分布为 $F(\xi)$ 的随机时间. 我们假定拥有非常多工人, 使得所有订货都不推延地予以处理. 令 $W(t)$ 表示在时刻 t 之前已发出订单但尚未完成的订单数. 试求

$$\lim_{t\to\infty}\Pr\{W(t)\leqslant k\}.$$

(b) 设 $V(t)$ 表示在时刻 0 没有未完成的订货情形下完成所有至时刻 t 的订货任务所需的时间. 试求 $V(t)$ 的概率分布, 即求

$$\Pr\{V(t)<y\}=F(y,t).$$

提示：首先假设第一次定货到达的时间，继而写出关于 $F(y,t)$ 的递推关系式.

27. 连续函数 $g(x), x \geqslant 0$ 的拉普拉斯变换为 $g^*(\theta) = \int_0^\infty \mathrm{e}^{-\theta x} g(x)\mathrm{d}x$, 其中 $\theta > 0$. 若一更新过程具有寿命密度函数 $f(x)$, $m(t) = \mathrm{d}M(t)/\mathrm{d}t$ 是其更新函数的导数. 试证明公式

$$m^*(\theta) = \frac{f^*(\theta)}{1 - f^*(\theta)},$$

若 (i) $f(x) = \lambda \mathrm{e}^{-\lambda x}, x \geqslant 0$; (ii) $f(x) = x\mathrm{e}^{-x}, x \geqslant 0$, 分别具体计算其 $m^*(\theta)$.

28. 设一更新过程具有均值 $(\sigma^2 + \mu^2)/2\mu$, 其中 σ^2 和 μ 分别是间距分布的方差和均值. 试求其现龄随机变量 δ_t 当 $t \to \infty$ 时的极限分布.

29. 设一更新过程间距分布的均值和方差分别为 μ 和 σ^2, δ_t 是其现龄随机变量. 试证明

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t \delta_\tau \mathrm{d}\tau = (\sigma^2 + \mu^2)/2\mu.$$

236

30. 若 $X_1, X_2, \cdots$ 是更新过程中间隔时间. 若 $\Pr\{X_k = 1\} = p$ 和 $\Pr\{X_k = 2\} = q = 1 - p$. 试证明

$$E[N_n] = \frac{n}{1+q} - \frac{q^2}{(1+q)^2} + \frac{q^{n+2}}{(1+q)^2}, \quad n = 2, 4, \cdots,$$

其中 N_n 是至时刻 (离散时间)n 为止的平均更新数.

参考书目

W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. II. Wiley, New York, 1966.

237

第6章　鞅

随机过程被它们的随机变量之间的依赖关系所刻划. 鞅所表示的关系大量出现在各种问题中, 它已成为理论的和应用的概率论基本工具, 可用于计算吸收概率, 分析连续时间过程的轨道结构, 导出有关随机过程的不等式, 分析序贯决策和控制模型以及大量其他用途.

鞅理论需要广泛使用条件期望, 因此建议读者复习列在第 1 章的有关条件期望的性质.

6.1　初步定义和例子

我们从鞅概念的最早形式开始, 它虽然已过时, 但仍然是有意义的.

定义 1.1　一个随机过程 $\{X_n; n=0,1,\cdots\}$ 是一个鞅, 如果对于 $n=0,1,\cdots$

(i) $E[|X_n|]<\infty$,

(ii) $E[X_{n+1}|X_0,\cdots,X_n]=X_n$.

设 X_n 是赌徒在第 n 次赌博时所拥有的财富. 鞅的性质体现 "公平赌博" 的概念. 在这种赌博中赌徒在下一次赌博将拥有的财富按平均来说与现在的财富是一样的, 不受过去情况的影响. 事实上, "鞅" 的名称来自一种赌博策略的法文首字母缩略词, 这种策略就是一直加倍赌注直到取胜为止. 现在, 鞅理论已应用于一般概率理论和数学分析的许多分支和多种领域. 若仅限于赌博来考虑鞅理论是不适当的, 并易使人误解.

关于鞅的更加一般和贴切的定义我们将在 6.7 节详细论述. 除非另外说明, 下面遇到的所有随机变量都假设为实值的.

238

定义 1.2　设 $\{X_n, n=0,1,\cdots\}$ 和 $\{Y_n, n=0,1,\cdots\}$ 是两个随机过程. 我们说 $\{X_n\}$ 是关于 $\{Y_n\}$ 的一个鞅, 如果对于 $n=0,1,\cdots$

(i) $E\{|X_n|\}<\infty$, (1.1)

(ii) $E\{X_{n+1}|Y_0,\cdots,Y_n\}=X_n$. (1.2)

把 $(Y_0,\cdots,Y_n)$ 看作至 n 为止的信息或历史通常是有用的. 这样, 在赌博的情况下, 这个历史所包含的信息可能要比仅由过去财富序列 $(X_0,\cdots,X_n)$ 所提供的信息更多一些, 例如赌徒不赌时的赌博结果. 实际上不需要限制 Y_k 是一个实的随机变量. 一般来说, 它可以是有限维甚至无限维的向量. 当特殊序列 $\{Y_n\}$ 不重要或比较显然时, 我们将不提它, 仅说 "$\{X_n\}$ 是一个鞅".

历史决定 X_n 是指: X_n 是 $Y_0, Y_1, \cdots, Y_n$ 的函数, 即 $Y_0, Y_1, \cdots, Y_n$ 值决定 X_n 的值. 由 (1.2) 知 X_n 是 $Y_0, Y_1, \cdots, Y_n$ 的特定函数:

$$X_n = E[X_{n+1}|Y_0, \cdots, Y_n]$$

由条件期望的性质

$$E[g(Y_0, \cdots, Y_n)|Y_0, \cdots, Y_n] = g(Y_0, \cdots, Y_n), \tag{1.3}$$

我们推得

$$E[X_n|Y_0, \cdots, Y_n] = X_n,$$

借助全概率公式, 得到

$$E[X_{n+1}] = E\{E[X_{n+1}|Y_0, \cdots, Y_n]\} = E[X_n],$$

再由归纳法知对所有的 n 有,

$$E[X_n] = E[X_0].$$

若干例子

这些例子表明鞅过程涉及面是非常广泛的. 我们先从一些重要具体情况开始, 然后讨论比较一般的情况.

239

(a) 独立随机变量之和

设 $Y_0 = 0$ 和 $Y_1, Y_2 \cdots$ 是独立随机变量序列, 且对所有 n, $E[|Y_n|] < \infty$ $E[Y_n] = 0$. 如果 $X_0 = 0$ 且 $X_n = Y_1 + \cdots + Y_n (n \geqslant 1)$, 则 $\{X_n\}$ 是关于 $\{Y_n\}$ 的一个鞅. 由

$$E[|X_n|] \leqslant E[|Y_1|] + \cdots + E[|Y_n|] < \infty$$

及

$$\begin{aligned} E[X_{n+1}|Y_0, \cdots, Y_n] &= E[X_n + Y_{n+1}|Y_0, \cdots, Y_n] \\ &= E[X_n|Y_0, \cdots, Y_n] + E[Y_{n+1}|Y_0, \cdots, Y_n] \\ &= X_n + E[Y_{n+1}] \quad (\text{由于}\{Y_i\}\text{的独立性}) \\ &= X_n \quad (\text{由于}E[Y_m] = 0), \end{aligned}$$

我们验证了 (1.1) 和 (1.2).

(b) 更一般的和

假设 $Z_i = g_i(Y_0, \cdots, Y_i)$, 其中 $\{Y_i\}$ 是任意随机变量序列, $\{g_i\}$ 是函数序列. 令 f 是满足如下条件的函数

$$E[|f(Z_k)|] < \infty, \quad k = 0, 1, \cdots.$$

令 a_k 是 k 个实变量的有界函数. 则

$$X_n = \sum_{k=0}^{n} \{f(Z_k) - E[f(Z_k)|Y_0, \cdots, Y_{k-1}]\} a_k(Y_0, \cdots, Y_{k-1})$$

定义了一个关于 $\{Y_n\}$ 的鞅. (这里约定, 当 $k = 0$ 时 $E[f(Z_k)|Y_0, \cdots, Y_{k-1}] = E[f(Z_k)]$.) 由于 a_k 是有界的, 可设对于所有的 $y_0, \cdots, y_{k-1}$

$$|a_k(y_0, \cdots, y_{k-1})| \leqslant A_k.$$

于是我们有

$$E[|X_n|] \leqslant 2 \sum_{k=0}^{n} A_k E[|f(Z_k)|] < \infty.$$

令 $B_k = \{f(Z_k) - E[f(Z_k)|Y_0, \cdots, Y_{k-1}]\} a_k(Y_0, \cdots, Y_{k-1})$. 则利用 (1.3), 得

$$E[B_k|Y_0, \cdots, Y_{k-1}] = 0.$$

这样,

$$\begin{aligned} E[X_n|Y_0, \cdots, Y_{n-1}] &= E[X_{n-1}|Y_0, \cdots, Y_{n-1}] + E[B_n|Y_0, \cdots, Y_{n-1}] \\ &= X_{n-1}, \end{aligned}$$

240 由此证明了鞅的性质.

(c) 和的方差

设 $Y_0 = 0$ 和 $Y_1, Y_2, \cdots$ 是独立同分布随机变量, $E[Y_k] = 0$ 且 $E[Y_k^2] = \sigma^2, k = 1, 2, \cdots$. 令 $X_0 = 0$ 及

$$X_n = \left(\sum_{k=1}^{n} Y_k\right)^2 - n\sigma^2.$$

于是 $E[|X_n|] \leqslant 2n\sigma^2 < \infty$, 且

$$\begin{aligned} E[X_{n+1}|Y_0, \cdots, Y_n] &= E\left[\left(Y_{n+1} + \sum_{k=1}^{n} Y_k\right)^2 - (n+1)\sigma^2 | Y_0, \cdots, Y_n\right] \\ &= E[Y_{n+1}^2 + 2Y_{n+1} \sum_{k=1}^{n} Y_k + \left(\sum_{k=1}^{n} Y_k\right)^2 - (n+1)\sigma^2 | Y_0, \cdots, Y_n] \\ &= X_n + E[Y_{n+1}^2|Y_0, \cdots, Y_n] + 2E[Y_{n+1}|Y_0, \cdots, Y_n]\left(\sum_{k=1}^{n} Y_k\right) - \sigma^2 \\ &= X_n. \end{aligned}$$

这样, $\{X_n\}$ 是一个关于 $\{Y_n\}$ 的鞅.

(d) 马尔可夫链的右正则序列及其所导出的鞅

有一个常规的但却高产的办法去找到与马尔可夫过程有关的鞅. 设 $Y_0, Y_1, \cdots$ 表示具有转移概率矩阵 $\boldsymbol{P} = \| p_{ij} \|$ 的马尔可夫链: 令 f 是关于 $\boldsymbol{P}$ 的有界右正则序列, 即 $f(i)$ 是非负的且满足

$$f(i) = \sum_j P_{ij} f(j). \tag{1.4}$$

(见第 2 卷第 11 章). 取 $X_n = f(Y_n)$, 因为 f 有界, 则 $E[|X_n|] < \infty$, 且

$$\begin{aligned} E[X_{n+1}|Y_0, \cdots, Y_n] &= E[f(Y_{n+1})|Y_0, \cdots, Y_n] \\ &= E[f(Y_{n+1})|Y_n] \quad (\text{由马尔可夫性}) \\ &= \sum_j P_{Y_n, j} f(j) \end{aligned}$$

(由于 $E[f(Y_{n+1})|Y_n = i] = \sum_j P_{ij} f(j)$)

$$\begin{aligned} &= f(Y_n) \quad (\text{根据 } (1.4)) \\ &= X_n. \end{aligned}$$

241

许多鞅初看起来似乎与此无关但实际上却以这种方式出现, 或以我们下面例子更一般的形式出现.

(e) 由转移矩阵特征向量导出的鞅

构成鞅的最广泛方法可能是由这个例子所叙述的方法 (见初等问题 8, 15,18, 19, 21 和 23). 设 $Y_0, Y_1, \cdots$ 是马尔可夫链, 具有转移概率矩阵 $\boldsymbol{P} = \| P_{ij} \|$. 向量 f 是 $\boldsymbol{P}$ 的右特征向量, 如果对某个特征值 λ 以及所有的 i 有

$$\lambda f(i) = \sum_j P_{ij} f(j),$$

如果 f 是 $\boldsymbol{P}$ 的右特征向量, 且对于所有的 $n, E[|f(Y_n)|] < \infty$, 则

$$X_n = \lambda^{-n} f(Y_n)$$

是一个鞅, 因为

$$\begin{aligned} E[X_{n+1}|Y_0, \cdots, Y_n] &= E[\lambda^{-n-1} f(Y_{n+1})|Y_0, \cdots, Y_n] \\ &= \lambda^{-n} \lambda^{-1} E[f(Y_{n+1})|Y_n] \\ &= \lambda^{-n} \lambda^{-1} \sum_j P_{Y_n, j} f(j) = \lambda^{-n} f(Y_n) \\ &= X_n. \end{aligned}$$

更一般地, 假设 $Y_0, Y_1, \cdots$ 是一个离散时间马尔可夫过程, 其转移分布函数是

$$F(y|z) = \Pr\{Y_{n+1} \leqslant y | Y_n = z\}. \tag{1.5}$$

如果对于所有的 n,

$$E[|f(Y_n)|] < \infty,$$

且对于所有的 y

$$\lambda f(y) = \int f(z) \mathrm{d}F(z|y),$$

则 $X_n = \lambda^{-n} f(Y_n)$ 是一个鞅.

后面的例子足够说明这个构成鞅的方法的广泛性.

(f) 分支过程

设 $\{Y_n\}$ 是一分支过程 (2.2 节例 F) 并假设后代分布的均值 $m < \infty$, 则 $X_n = m^{-n} Y_n$ 是一个鞅. 为了证实此结论, 我们用 $Z^n(j)$ 表示第 n 代中由第 j 个母体产生的后代数目. 则

$$Y_{n+1} = Z^n(1) + \cdots + Z^{(n)}(Y_n),$$

242 其中 $Z^{(n)}(i), i = 1, 2, \cdots, Y_n$ 是独立同分布的. 显然

$$E[Y_{n+1}|Y_n] = Y_n E[Z^{(n)}(1)] = mY_n,$$

所以 m 是函数 $f(y) = y$ 的特征值. 由此推得 $X_n = m^{-n} Y_n$ 是一个鞅.

(g) Wald 鞅

令 $Y_0 = 0$ 并设 $Y_1, Y_2, \cdots$ 是独立同分布随机变量, 且对某个 $\lambda \neq 0$, 其矩母函数$\phi(\lambda) = E[\exp\{\lambda Y_k\}]$ 有限. 那么, $X_0 = 1$ 和 $X_n = \phi(\lambda)^{-n} \exp\{\lambda(Y_1 + \cdots + Y_n)\}$ 确定了一个关于 $\{Y_n\}$ 的鞅, 因为函数 $f(y) = \mathrm{e}^{\lambda y}$ 是部分和 $S_n = Y_1 + \cdots + Y_n$ 马尔可夫过程的一个特征函数, 其相应的特征值是 $\phi(\lambda)$. 事实上, 和 (1.5) 一致, 在这种情况下转移分布函数是

$$\mathrm{P_r}\{S_{n+1} \leqslant y | S_n = x\} = G(y - x).$$

这里 G 是 Y_k 的公共分布函数.

现在我们来计算 $E[f(S_{n+1})|S_n = x]$, 利用变量替换, 得

$$\int \mathrm{e}^{\lambda y} \mathrm{d}_y G(y - x) = \mathrm{e}^{\lambda x} \int \mathrm{e}^{\lambda \xi} \mathrm{d}G(\xi) = \mathrm{e}^{\lambda x} \phi(\lambda).$$

这个恒等式显然证实了前面的结论.

作为实例, 假设 $Y_1, Y_2, \cdots$ 是独立的正态分布, 其均值为 0, 方差为 σ^2, 则

$$\phi(\lambda) = E[\exp\{\lambda Y_1\}] = \exp\left\{\frac{1}{2}\lambda^2\sigma^2\right\}$$

和

$$X_n = \exp\left\{\lambda(Y_1 + \cdots + Y_n) - \frac{n}{2}\lambda^2\sigma^2\right\}.$$

选取 $\lambda = \mu/\sigma^2$, 其中 μ 是一个任意常数, 我们得到

$$X_n = \exp\left\{\frac{\mu}{\sigma^2}(Y_1 + \cdots + Y_n) - \frac{n\mu^2}{2\sigma^2}\right\}.$$

这个鞅还将在例 (j) 中出现.

(h) 特征向量方法的推广

设 $Y_0, Y_1, \cdots$ 是任一随机变量序列, 却具有有限的一阶绝对均值 $E[|Y_n|] < \infty$. 假设, 对 $n = 0, 1, 2, \cdots$,

$$E[Y_{n+1}|Y_0, \cdots, Y_n] = a_n + b_n Y_n, \quad b_n \neq 0. \tag{1.6}$$

243

令 $l_{n+1}(z)$ 是线性函数 $l_{n+1}(z) = a_n + b_n z$, 它的逆是 $l_{n+1}^{-1}(y) = (y - a_n)/b_n$, 并且令 $L_n(y) = l_1^{-1}(l_2^{-1}(\cdots(l_n^{-1}(y))\cdots))$. 则对于任意常数 k, $X_n = kL_n(Y_n)$ 是一个鞅, 因为

$$\begin{aligned}\frac{1}{k}E[X_{n+1}|Y_0, \cdots, Y_n] &= E[L_{n+1}(Y_{n+1})|Y_0, \cdots, Y_n] \\ &= L_{n+1}\{E[Y_{n+1}|Y_0, \cdots, Y_n]\}\end{aligned}$$

(因为 L_{n+1} 是它的自变量的线性函数)

$$\begin{aligned}&= L_{n+1}(l_{n+1}(Y_n)), \quad (\text{由式 } (1.6)) \\ &= L_n(Y_n) = \frac{1}{k}X_n.\end{aligned}$$

具体地, 设 Y_0 是 $[0, 1]$ 上均匀分布, 若已知 Y_n, Y_{n+1} 是在 $[Y_n, 1]$ 上均匀分布. 则 $X_n = 2^n[1 - Y_n]$ 是一个鞅. 我们直接检验如下:

$$\begin{aligned}E[X_{n+1}|Y_0, \cdots, Y_n] &= 2^{n+1}[1 - E[Y_{n+1}|Y_n]] \\ &= 2^{n+1}\left[1 - \frac{1}{2}(1 + Y_n)\right] \\ &= 2^n(1 - Y_n) = X_n.\end{aligned}$$

(i) 罐模型

这是 (h) 的另外一个例子. 这个模型已应用于增长过程的研究.

考虑一个罐, 在第 0 步包含一个红球和一个绿球. 随机地从罐中抽出一个球, 然后把它和另外一个相同颜色的球又一并放回罐中. 这个试验步骤无限次重复进行. 令 X_n 是第 n 步红球所占比例, 于是 $Y_n=(n+2)X_n$ 是红球的数目. 那么, $\{X_n\}$ 是一个关于 $\{Y_n\}$ 的鞅, 若已知 $Y_n=k$, 我们有

$$Y_{n+1}=\begin{cases} k+1, & \text{以概率} k/(n+2), \\ k, & \text{以概率} 1-k/(n+2). \end{cases}$$

因此

$$E[Y_{n+1}|Y_n=k]=\frac{(k+1)k+k(n+2-k)}{n+2}=k(n+3)/(n+2).$$

即

244

$$E[Y_{n+1}|Y_n]=b_nY_n.$$

其中 $b_n=(n+3)/(n+2)$. 然后, 利用 (h) 中的记号, $l_n(z)=b_{n-1}z, l_n^{-1}(y)=y/b_{n-1}$, 以及

$$\begin{aligned} L_n(y) &= \frac{y}{b_0b_1\cdots b_{n-1}} \\ &= \frac{2}{3}\times\frac{3}{4}\times\cdots\times\frac{n+1}{n+2}y \\ &= \frac{2}{n+2}y. \end{aligned}$$

于是

$$X_n=\frac{1}{2}L_n(Y_n)=\frac{1}{n+2}Y_n$$

是一个鞅.

(j) 似然比

设 $Y_0, Y_1, \cdots$ 是独立同分布随机变量, 并令 f_0 和 f_1 是概率密度函数. 在统计假设检验理论中一个很重要的随机过程是似然比序列

$$X_n=\frac{f_1(Y_0)f_1(Y_1)\cdots f_1(Y_n)}{f_0(Y_0)f_0(Y_1)\cdots f_0(Y_n)},\quad n=0,1,\cdots.$$

为保证定义有意义, 假设对任意 $y, f_0(y)>0$. 由于 $Y_0, Y_1, \cdots$ 是独立的,

$$\begin{aligned} E[X_{n+1}|Y_0,\cdots,Y_n] &= E\Big[X_n\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}|Y_0,\cdots,Y_n\Big] \\ &= X_nE\Big[\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}\Big]. \end{aligned}$$

当 Y_k 的共同分布具有概率密度函数 f_0 时, $\{X_n\}$ 是一个关于 $\{Y_n\}$ 的鞅. 为了证实这一点, 我们仅需要验证

$$E\left[\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}\right]=1.$$

但,

$$E\left[\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}\right]=\int\left\{\frac{f_1(y)}{f_0(y)}\right\}f_0(y)\mathrm{d}y=\int f_1(y)\mathrm{d}y=1,$$

此即所要证明的.

245

作为一个例子, 假设 f_0 是正态密度函数, 均值为 0, 方差为 σ^2, 而 f_1 是均值为 μ 和方差为 σ^2 的正态函数, 则

$$\frac{f_1(y)}{f_0(y)}=\exp\left\{\frac{2\mu y-\mu^2}{2\sigma^2}\right\},$$

并且

$$X_n=\exp\left\{\frac{\mu}{\sigma^2}(Y_1+\cdots+Y_n)-\frac{n\mu^2}{2\sigma^2}\right\}.$$

这个鞅早已出现在例 (g) 中.

由似然比构造出来的鞅在研究假设检验序贯过程的性质中有很多的应用.

(k) Doob 鞅

设 $Y_0,Y_1,\cdots$ 是任意一个随机变量序列, 且 X 是满足 $E[|X|]<\infty$ 的随机变量, 则

$$X_n=E[X|Y_0,\cdots,Y_n]$$

构成一个关于 $\{Y_n\}$ 的鞅, 称为 Doob 过程. 首先

$$\begin{aligned}E[|X_n|]&=E\{|E[X|Y_0,\cdots,Y_n]|\}\\&\leqslant E\{E[|X||Y_0,\cdots,Y_n]\}=E[|X|]<\infty.\end{aligned}$$

其次, 由条件期望的全概率公式[1]有

$$\begin{aligned}E[X_{n+1}|Y_0,\cdots,Y_n]&=E\{E[X|Y_0,\cdots,Y_{n+1}]|Y_0,\cdots,Y_n\}\\&=E[X|Y_0,\cdots,Y_n]=X_n.\end{aligned}$$

1. 这里将通常的条件期望的全概率公式推广为公式 $E[X|Z]=E\{E[X|Y,Z]|Z\}$, 其中随机变量 X 满足条件 $E[|X|]<\infty$. 读者可补充其证明.

(1) 拉东 – 尼古丁导数.

假设 Z 是在 $[0,1)$ 上均匀分布的随机变量, 定义一个新的随机变量 Y_n:

$$Y_n = k/2^n,$$

其中 k(取决于 n 和 Z) 满足

246

$$\frac{k}{2^n} \leqslant Z < \frac{k+1}{2^n}.$$

请注意, 当 n 增加时 Y_n 如何不断提供更多的关于 Z 的信息. 事实上, $Y_n \leqslant Z < Y_n + \left(\frac{1}{2}\right)^n$, 从而 Y_n 确定了 Z 的有尽二进位分解中的前 n 位数字.

令 f 是 [0, 1] 上有界函数, 构选差商

$$X_n = 2^n\{f(Y_n + 2^{-n}) - f(Y_n)\}.$$

我们断定 $\{X_n\}$ 是一个关于 $\{Y_n\}$ 的鞅. 首先注意到, 在条件 $Y_0, \cdots, Y_n \cdots$ 下,Z 在 $[Y_n, Y_n + 2^{-n}]$ 上均匀分布, 于是 Y_{n+1} 等可能地为 Y_n 或 $Y_n + 2^{-(n+1)}$, 故

$$\begin{aligned} &E[X_{n+1}|Y_0, \cdots, Y_n] = 2^{n+1}E[f(Y_{n+1} + 2^{-(n+1)}) - f(Y_{n+1})|Y_0, \cdots, Y_n] \\ &= 2^{n+1}\left\{\frac{1}{2}[f(Y_n + 2^{-(n+1)}) - f(Y_n)] + \frac{1}{2}[f(Y_n + 2^{-n}) - f(Y_n + 2^{-(n+1)})]\right\} \\ &= 2^n\left\{f(Y_n + 2^{-n}) - f(Y_n)\right\} = X_n. \end{aligned}$$

注意到

$$X_n = \frac{f(Y_n + 2^{-n}) - f(Y_n)}{2^{-n}}$$

近似地是 f 在 Z 的导数. 事实上, 在十分一般的条件下, 可以证明序列 $\{X_n\}$ 以概率 1 收敛于某个随机变量 $X_\infty = X_\infty(Z)$, 并且 $X_n = E[X_\infty|Y_0, \cdots, Y_n]$(见 6.7 节末), 随机变量 $X_\infty(Z)$ 称为 f 在 Z 的拉东 – 尼古丁导数. 这样, 鞅的性质与函数的微分理论甚至与数学分析许多其他方面都有着密切联系.

结果提要

下一节将处理比较一般的情况, 在那里鞅的等式将被不等式所代替. 然后我们介绍鞅理论两个主要结果, **可选抽样定理和鞅收敛定理**, 以及这些定理的诸多应用.

可选抽样定理告诉我们, 在十分一般的情况下, 只要 X_n 是一个鞅, 则 $X_{T_n} = Z_n$ 也构成一个鞅, 其中 $\{T_n\}$ 是随机选择时间集合, 它们构成一个 "马尔可夫时间" 的递增序列. 一个马尔可夫时间 T 具有如下性质: 事件 $\{T = n\}$ 仅仅由直至 n 为止的

历史 $(Y_0, Y_1, \cdots, Y_n)$ 所确定. 可选抽样 (或停止) 定理在序贯决策问题中. 在推导某些不等式以及在计算与某些随机过程有关的各种事件的概率中都有着很多的应用. 247

鞅的收敛定理给出当 $n \to \infty$ 时使得鞅 X_n 收敛于极限随机变量 X_∞ 的一般条件. 这些定理在分析过程的轨道构造以及确定多种类型随机过程的各种函数的渐近分布时具有重要意义.

6.2 上鞅和下鞅

为了多种目的, 需要一个由不等式建立的更一般的概念.

定义 2.1 设 $\{X_n, n = 0, 1, \cdots\}$ 和 $\{Y_n, n = 0, 1, \cdots\}$ 是随机过程. $\{X_n\}$ 称为一个关于 $\{Y_n\}$ 的上鞅, 如果对任意 n

(i) $E[X_n^-] > -\infty$, 其中 $x^- = \min\{x, 0\}$,

(ii) $E[X_{n+1}|Y_0, \cdots, Y_n] \leqslant X_n$, (2.1)

(iii) X_n 是 $(Y_0, \cdots, Y_n)$ 的函数.

我们称 $\{X_n\}$ 是一个关于 $\{Y_n\}$ 的下鞅, 如果对任意 n 有,

(i) $E[X_n^+] < \infty$, 其中, $x^+ = \max\{0, x\}$,

(ii) $E[X_{n+1}|Y_0, \cdots, Y_n] \geqslant X_n$, (2.2)

(iii) X_n 是 $(Y_0, \cdots, Y_n)$ 的函数.

正如鞅的情况那样, 我们经常省略 $\{Y_n\}$, 假如它不太重要或从上下文来看是显然的.

上述定义中的第三个条件指出了 X_n 由到时刻 n 为止的历史所确定, 或等价地, 至 n 为止的可得到的信息能够确定 X_n 的值. 前面已指出, 对于鞅而言, 如此确定方式自动满足, 因为 X_n 作为 $\{Y_i\}_{i=0}^n$ 的函数有等式

$$X_n = E[X_{n+1}|Y_0, \cdots, Y_n]. \tag{2.3}$$

而对上、下鞅而言, 鞅的等号被不等号所代替, 因此, X_n 能被 $(Y_0, Y_1, \cdots, Y_n)$ 所确定这个必要条件应当明确地加上. 我们有时通过下面函数的表达式来表示这个函数关系:

$$X_n = X_n(Y_0, \cdots, Y_n).$$

248

注意, $\{X_n\}$ 是一个关于 $\{Y_n\}$ 的上鞅当且仅当 $\{-X_n\}$ 是下鞅. 类似地, $\{X_n\}$ 是一个关于 $\{Y_n\}$ 的鞅当且仅当 $\{X_n\}$ 既是下鞅又是上鞅. 因此, 关于上鞅的命题能被改为关于鞅和下鞅的相应命题. 为了避免重复书写, 我们通常仅给出三种情况中的一种情况的证明.

例 设 $\{Y_n\}$ 是马尔可夫链, 具有转移概率矩阵 $\boldsymbol{P} = \| P_{ij} \|$. 如果 f 是 $\boldsymbol{P}$ 的右上正则序列 (即对任意 i 满足 $\sum_j P_{ij} f(j) \leqslant f(i)$ 的非负序列), 则 $X_n = f(Y_n)$ 定义了一个关于 $\{Y_n\}$ 的上鞅. 这个论断的证明可仿照 6.1 节例 (d) 进行.

当然, 在下鞅和右下正则序列 (即对任意 i 满足 $f(i) \leqslant \sum_j P_{ij}f(j)$ 的非负序列) 之间存在类似的关系, 倘若我们假设 $E[f(Y_n)] < \infty$.

Jensen 不等式 称定义在区间 I 的函数 ϕ 是**凸函数**, 如果对任意 $x_1, x_2 \in I$ 和 $0 < \alpha < 1$, 有

$$\alpha\phi(x_1) + (1-\alpha)\phi(x_2) \geqslant \phi(\alpha x_1 + (1-\alpha)x_2). \tag{2.4}$$

由 (2.4) 开始直接使用归纳法得

$$\sum_{i=1}^{m} \alpha_i \phi(x_i) \geqslant \phi\Big(\sum_{i=1}^{m} \alpha_i x_i\Big), \tag{2.5}$$

对所有 $x_1, x_2, \cdots, x_m \in I$ 和 $\alpha_i \geqslant 0, \sum_{i=1}^{m} \alpha_i = 1$ 成立. 如果 ϕ 是二次可微的, 则 ϕ 是凸的当且仅当对所有 x, $\dfrac{\mathrm{d}^2\phi}{\mathrm{d}x^2} \geqslant 0$. 这样, 凸性通常很容易验证. 如果 X 是随机变量, 它分别以概率 $\alpha_i (i = 1, 2, \cdots, m)$ 取值 x_i, 则方程 (2.5) 能简写为

$$E[\phi(X)] \geqslant \phi(E[X]). \tag{2.6}$$

Jensen 不等式指出, 当 ϕ 是 $(-\infty, +\infty)$ 上的凸函数时, (2.6) 对所有的实随机变量 X 都成立. 不等式 (2.6) 可视为 (2.5) 的连续积分形式. 相同的结论对条件期望也成立. 如果 ϕ 是凸的, 我们有

$$E[\phi(X)|Y_0, \cdots, Y_n] \geqslant \phi(E[X|Y_0, \cdots, Y_n]). \tag{2.7}$$

249

由上述事实我们可以用鞅来构造下鞅.

引理 2.1 设 $\{X_n\}$ 是一个关于 $\{Y_n\}$ 的鞅. 如果 ϕ 是一个凸函数, 且对任意 $n, E[\phi(X_n)^+] < \infty$, 则 $\{\phi(X_n)\}$ 是一个关于 $\{Y_n\}$ 的下鞅. 特别地, $\{|X_n|\}$ 通常是一个下鞅. 倘若对任意 n 有 $E[X_n^2] < \infty$, 则 $\{X_n^2\}$ 是一个下鞅.

证明 我们只需证明下鞅不等式, 其他性质是比较容易证明的. 利用 Jensen 不等式, 我们有

$$\begin{aligned} E[\phi(X_{n+1})|Y_0, \cdots, Y_n] &\geqslant \phi(E[X_{n+1}|Y_0, \cdots, Y_n]) \\ &= \phi(X_n). \end{aligned}$$

■

下面是类似结果, 证明从略.

引理 2.2 设 $\{X_n\}$ 是一个关于 $\{Y_n\}$ 的下鞅. 如果 ϕ 是凸的增函数, 那么 $\phi(X_n)$ 是一个下鞅, 倘若 $E[\phi(X_n)^+] < \infty$.

(注意, 对 $\{X_n\}$ 的要求条件比较弱, 仅仅是下鞅, 但对 ϕ 的要求比较高, ϕ 是增的凸函数.)

由此, 如果 $\{X_n\}$ 是下鞅, 且

$$\widetilde{X}_n = \begin{cases} X_n, & 若X_n > -c, \\ -c, & 若X_n \leqslant -c, \end{cases} \tag{2.8}$$

其中 c 是固定的, 则 $\{\widetilde{X}_n\}$ 是下鞅, 且对任意 $n, E[|\widetilde{X}_n|] < \infty$. 特别地, 当 $\{X_n\}$ 是下鞅时, $\{X_n^+\}$ 也是下鞅.

初等性质

我们在一个命题中包含上鞅和鞅的结果, 关于上鞅的假设和结论写在括号里. (把 $\{X_n\}$ 改为 $\{-X_n\}$, 可以导出关于下鞅的相应结果.)

(a) 如果 $\{X_n\}$ 是一个关于 $\{Y_n\}$ 的 (上) 鞅, 则对每个 $k \geqslant 0$ 成立

$$E[X_{n+k}|Y_0, \cdots, Y_n](\leqslant) = X_n. \tag{2.9}$$

证明 我们应用归纳法. 由定义, (2.9) 对于 $k=1$ 是正确的. 假设 (2.9) 对于 k 正确, 则

$$\begin{aligned} E[X_{n+k+1}|Y_0, \cdots, Y_n] &= E\{E[X_{n+k+1}|Y_0, \cdots, Y_n, \cdots, Y_{n+k}]|Y_0, \cdots, Y_n\} \\ (\leqslant) &= E\{X_{n+k1}|Y_0, \cdots, Y_n\} \\ (\leqslant) &= X_n. \end{aligned}$$

250

(b) 如果 $\{X_n\}$ 是 (上) 鞅, 则对 $0 \leqslant k \leqslant n$

$$E[X_n](\leqslant) = E[X_k](\leqslant) = E[X_0]. \tag{2.10}$$

证明 利用 (2.9), 我们对式

$$E[X_n|Y_0, \cdots, Y_k](\leqslant) = X_k$$

取期望, 得

$$E[X_n] = E\{E[X_n|Y_0, \cdots, Y_k]\}(\leqslant) = E[X_k].$$

同理可证 $E[X_k](\leqslant) = E[X_0]$. ■

(c) 假设 $[X_n]$ 是一个关于 $\{Y_n\}$ 的 (上) 鞅并且 g 是 $Y_0, \cdots, Y_n$ 的 (非负) 函数, 倘若其期望存在, 则

$$E[g(Y_0, \cdots, Y_n)X_{n+k}|Y_0, \cdots, Y_n](\leqslant) = g(Y_0, \cdots, Y_n)X_n. \tag{2.11}$$

证明 由于 $g(Y_0,\cdots,Y_n)$ 是由 $(Y_0,\cdots,Y_n)$ 决定的, 利用条件期望的基本性质, 我们有

$$E[g(Y_0,\cdots,Y_n)X_{n+k}|Y_0,\cdots,Y_n]=g(Y_0,\cdots,Y_n)E[X_{n+k}|Y_0,\cdots,Y_n]$$
$$(\leqslant)=g(Y_0,\cdots,Y_n)X_n.$$

(对于上鞅情况 $(\leqslant)$, 我们需要 $g\geqslant 0$.) ■

序贯决策模型

考虑具有有限数 S 个状态的系统, 状态用正整数 $1,2,\cdots,S$ 表示. 我们定期地 (譬如每天一次) 观察系统的现时状态, 然后从 A 个可能行动 $1,2,\cdots,A$ 中选择一个, 作为现时状态 s 和所选择行动 a 的联合结果, 下面两个事件发生: (i) 我们得到直接收入 $i(s,a)$; (ii) 系统转移到新的状态, 转移到新状态 s' 的概率由已知函数 $q=q(s'|s,a)$ 确定. 我们的问题是确定选择行动的策略使得至 N 个周期时间为止总的平均收入最大.

设 $S_0,A_0,S_1,\cdots,A_{N-2},S_{n-1},A_{N-1}$ 是状态和行动的交错序列. 一个策略 π 是一组函数 $\pi_0,\cdots,\pi_{N-1}$, 其中 π_n 是规定行动 A_n 作为观察到的历史 $S_0,A_0,\cdots,A_{n-1}$, S_n 的函数, 即, 如果策略 π 被采用, 则

251

$$A_n=\pi_n(S_0,A_0,\cdots,A_{n-1},S_n).$$

在策略 π 之下, 作为初始状态 $S_0=s$ 函数的平均收入为

$$I(\pi,s)=E\left[\sum_{k=0}^{N-1}i(S_k,A_k)\right].$$

我们希望选择 π 使得 $I(\pi,s)$ 最大.

用向后递推的方法定义函数 $f_0,\cdots,f_N$. 令

$$f_N(s)=0,\qquad \text{对所有 } s, \tag{2.12}$$

并对 $n=1,2,\cdots,N$,

$$f_{n-1}(s)=\max_a\left\{i(s,a)+\sum_{s'}f_n(s')q(s'|s,a)\right\}. \tag{2.13}$$

则依据 6.1 节例 (b)

$$X_n=\sum_{k=1}^{n}\{f_k(S_k)-E[f_k(S_k)|S_0,A_0,\cdots,S_{k-1},A_{k-1}]\}$$

关于 $\{Y_n\}=\{(S_n,A_n)\}$ 构成一个鞅. 显然,

$$E[X_n]=0. \tag{2.14}$$

由 (2.13) 知

$f_{n-1}(s)\geqslant i(s,a)+\sum_{s'}f_n(s')q(s'|s,a)$, 对所有的 s 和 a, 所以, 特别地,

$$\begin{aligned}f_{k-1}(S_{k-1})&\geqslant i(S_{k-1},A_{k-1})+\sum_{s'}f_k(s')q(s'|S_{k-1},A_{k-1})\\&=i(S_{k-1},A_{k-1})+E[f_k(S_k)|S_0,A_0,\cdots,S_{k-1},A_{k-1}],\end{aligned} \tag{2.15}$$

不管采用什么策略上式均成立. 代入 (2.14) 得到

$$\begin{aligned}0=E[X_N]&=E\left[\sum_{k=1}^{N}\{f_k(S_k)-E[f_k(S_k)|S_0,A_0,\cdots,S_{k-1},A_{k-1}]\}\right]\\&\geqslant E\Big[\sum_{k=1}^{N}\{f_k(S_k)+i(S_{k-1},A_{k-1})-f_{k-1}(S_{k-1})\}\Big]\\&=E\Big[\sum_{k=0}^{N-1}i(S_k,A_k)+f_N(S_N)-f_0(S_0)\Big].\end{aligned}$$

即

$$E[f_0(S_0)]\geqslant E\Big[\sum_{k=0}^{N-1}i(S_k,A_k)\Big].$$

252

如果 $S_0=s$, 则对任何策略 π

$$f_0(s)\geqslant I(\pi,s).$$

这就是说没有一个策略能够使得平均收入超过 $f_0(s)$. 这样, 如果我们有一个策略 π^* 满足

$$f_0(s)=I(\pi^*,s),$$

则这个策略显然是最优的, 对每个 s, 令

$$\pi^*_{n-1}(S_0,A_0,\cdots,A_{n-2},s)$$

是使 (2.13) 右边部分取到最大值的行动, 即如果 $A^*_{k-1}=\pi^*_{k-1}(S_0,A^*_0,\cdots,S_{k-1},)$, 则 (2.15) 化为等式

$$f_{k-1}(S_{k-1})=i(S_{k-1},A^*_{k-1})+E[f_k(S_k)|S_0,A^*_0,\cdots,S_{k-1},A^*_{k-1}].$$

如前述同样论证, 我们进一步得到等式

$$E[f_0(S_0)] = E\Big[\sum_{k=0}^{N-1} i(S_k, A_k^*)\Big]$$

或

$$f_0(s) = I(\pi^*, s).$$

这样的 π^*, 即在第 $n-1$ 步处于状态 $S_{n-1} = s$ 之下选择行动使得 (2.13) 的右边部分达到最大的策略是最优的.

6.3 可选抽样定理

考虑一个公平赌博, 第一盘赌博以相同概率赢取或输掉一元钱. 令 $Y_1, Y_2, \cdots$ 是独立同分布随机变量, 且 $\Pr\{Y_k = +1\} = \Pr\{Y_k = -1\} = \frac{1}{2}$. 设 $X_n = Y_1 + \cdots + Y_n$ 是赌博某一方至第 n 步净营利. 显然 $E[X_n] = 0$, 即净营利的平均值为 0. 但赌博者不需要一直赌下去, 也不需要预先指定一个特定时间停止. 他可以按照赌博进展的情况来确定何时停止赌博. 例如他可以在赢钱时停止赌博.

设 T 是赌博者停止赌博的时间, X_T 是当时的净营利. 我们已经知道对任意 n 有 $E[X_n] = 0$. 是否也一定有 $E[X_T] = 0$? 赌博者是否能够在赢时停止? 回答是肯定的, 但有一些限制条件.

首先, 由于我们不是先知先觉, 我们要求选择停止的时间仅仅取决于迄今为止所观测到的信息. 即, 我们要求赌博者停止在第 n 步这个事件仅仅取决于 $Y_0, \cdots, Y_n$. 随机变量 T 若对每一个 n 满足这个要求便称为**马尔可夫时间**, 有时也称为**停时**, 或**不依赖于将来的随机变量**(关于 $\{Y_n\}$). 随后, 我们将给出它的清晰定义.

253

即使在这个限制之下, 我们能够使 $E[X_T] > 0$! 例如 $T = \min\{n : X_n = 1\}$ 是马尔可夫时间, 因为 $\{T = n\}$ 当且仅当 $Y_1 + \cdots + Y_k < 1 \quad (k < n)$ 和 $Y_1 + \cdots + Y_n = 1$. 由于随机游动 $\{X_n\}$ 是常返的, $T < \infty$, 并且 $X_T \equiv 1$, 所以, $E[X_T] = 1 > 0$.

本节的目的是更进一步地讨论这个问题. 我们将说明上述定义的马尔可夫时间 T 具有某些局限性, 并且在实际中不一定可实现. 例如, 在上例中当 T 的均值为无限时, 在停止之前按平均有无限多的财富将失去, 所以赌博者要成功地采用这个策略必须有无限多财富.

另一方面, 在十分一般的条件下, 当 T 是具有有限期望的马尔可夫时间时, 对于一个鞅成立 $E[X_T] = E[X_0]$, 这个结果有广泛的应用, 其意义远远超出了赌博的范围.

马尔可夫时间

马尔可夫时间出现在很多地方. 这里有一个应用于马尔可夫链的例子. 假设 i 是马尔可夫链 $\{Y_n\}$ 的常返状态. 我们要证明

$$\Pr\{Y_n\text{回到}i\text{至少两次}\} = 1.$$

既然 i 是常返的, 我们知道

$$\Pr\{Y_n\text{回到}i\text{至少一次}\} = \Pr\{T_i < \infty\} = 1.$$

其中 $T_i = \min\{n \geqslant 1 : Y_n = i\}$ 是第一次回到 i 的时间, 利用马尔可夫性, 有

$$\begin{aligned}
&\Pr\{\text{回到}i\text{至少两次}\}\\
&= \Pr\{T_i < \infty\}\Pr\{\text{在}T_i\text{之后回到}i|T_i < \infty\}\\
&= \Pr\{T_i < \infty\}\Pr\{T_i < \infty|Y_0 = i\}\\
&= 1 \times 1 = 1.
\end{aligned}$$

这里有些问题需要进一步讨论. 如, 在 T_i 之后回到 i 的概率为什么和任何时刻回到 i 的概率相同? 下面给出的更详细的证明将说明何处利用了马尔可夫性以及随机时间 T_i 的什么性质使得上述成为可能. 我们注意到

254

$$\begin{aligned}
&\Pr\{\text{在}T_i\text{之后回到 }i|T_i = k\}\\
&= \Pr\{Y_{k+n} = i\text{对某些 }n = 1, 2, \cdots|Y_j \neq i, j = 1, 2, \cdots, k-1; Y_k = i\}\\
&= \Pr\{Y_{k+n} = i\text{对某些 }n = 1, 2, \cdots|Y_k = i\}(\text{由马尔可夫性})\\
&= \Pr\{Y_n = i\text{对某些 }n = 1, 2, \cdots|Y_0 = i\}(\text{利用平稳转移概率的性质})\\
&= \Pr\{T_i < \infty|Y_0 = i\} = 1.
\end{aligned}$$

这样,

$$\begin{aligned}
&\Pr\{T_i < \infty \text{ 且在 } T_i \text{ 之后回到 } i\}\\
&= \sum_{k=1}^{\infty} \Pr\{\text{在 } T_i \text{ 之后回到 } i|T_i = k\}\Pr\{T_i = k\}\\
&= \sum_{k=1}^{\infty} 1 \times \Pr\{T_i = k\} = 1.
\end{aligned}$$

这里的关键是事件 $\{T_i = k\}$ 与事件 $\{Y_j \neq i$, 对于 $j = 1, 2, \cdots, k-1; Y_k = i\}$ 相同, 并且仅依赖于 $(Y_0, \cdots, Y_k)$.

定义 3.1 随机变量 T 称为是一个关于 $\{Y_n\}$ 的马尔可夫时间, 如果 T 取值于 $\{0, 1, \cdots, \infty\}$, 并且对于每个 $n = 0, 1, \cdots$, 事件 $\{T = n\}$ 由 $(Y_0, \cdots, Y_n)$ 确定.

"确定" 意指 $\{T=n\}$ 的示性函数可视为 $Y_0,\cdots,Y_n$ 的函数, 即, 我们能够由过程 $Y_0,Y_1,\cdots,Y_n$ 至 n 为止取值的情况确定究竟 $T=n$ 或 $T\neq n$. 用式子将上述写为

$$\begin{aligned}I_{\{T=n\}}&=I_{\{T=n\}}(Y_0,\cdots,Y_n)\\&=\begin{cases}1, & 若\ T=n,\\0, & 若\ T\neq n.\end{cases}\end{aligned}$$

我们通常省去 $\{Y_n\}$, 仅仅说 "T 是马尔可夫时间". 如果 T 是一个马尔可夫时间, 则对于每个 n, 事件 $\{T\leqslant n\},\{T>n\},\{T\geqslant n\}$ 和 $\{T<n\}$ 也被 $(Y_0,\cdots,Y_n)$ 所确定. 事实上, 我们有

$$\begin{aligned}I_{\{T\leqslant n\}}&=\sum_{k=0}^{n}I_{\{T=k\}}(Y_0,\cdots,Y_k),\\I_{\{T>n\}}&=1-I_{\{T\leqslant n\}}(Y_0,\cdots,Y_n),\end{aligned}$$

等等.

反之, 如果对每个 n, 事件 $\{T\leqslant n\}$ 由 $(Y_0,\cdots,Y_n)$ 所确定, 则 T 是一个马尔可夫时间. 或者, 若对每个 n, 事件 $\{T\geqslant n\}$ 由 $(Y_0,\cdots,Y_n)$ 所确定, 则 T 也是一个马尔可夫时间. (见问题 20).

255

如果 $\{X_n\}$ 是一个关于 $\{Y_n\}$ 的鞅, 则对于每个 n, X_n 由 $(Y_0,\cdots,Y_n)$ 确定. 由此推得每个关于 $\{X_n\}$ 的马尔可夫时间同样也是一个关于 $\{Y_n\}$ 的马尔可夫时间. 当然, 同样命题对于上鞅和下鞅也是成立的.

马尔可夫时间的一些例子

(a) 固定时间 (即常数)$T\equiv k$ 是马尔可夫时间. 对所有 $Y_0,Y_1,\cdots$, 我们有

$$I_{\{T=n\}}(Y_0,\cdots,Y_n)=\begin{cases}0, & 如果\ n\neq k,\\1, & 如果\ n=k.\end{cases}$$

(b) 过程 $Y_0,Y_1,\cdots$ 首达状态空间子集 A 的时间也是马尔可夫时间, 即对

$$T(A)=\min\{n:Y_n\in A\},$$

我们有

$$I_{\{T(A)=n\}}\{Y_0,\cdots,Y_n\}=\begin{cases}1, & 如果\ Y_j\notin A, j=0,\cdots,n-1, Y_n\in A.\\0, & 其他.\end{cases}$$

(c) 更一般地, 对任何固定 k, 过程第 k 次到达集合 A 的时间也是一个马尔可夫时间. 然而, 一个过程到达一个集合的最后时间不是马尔可夫时间. 这是由于, 要确定某个特定的到达时间是否是最后时刻, 我们必须知道整个将来情况.

初等性质

(a) 如果 S 和 T 是马尔可夫时间, 则 $S+T$ 也是马尔可夫时间. 我们有

$$I_{\{S+T=n\}}=\sum_{k=0}^{n}I_{\{S=k\}}I_{\{T=n-k\}}.$$

(b) 两个马尔可夫时间中的较小者, 表示为

$$S\wedge T=\min\{S,T\},$$

也是马尔可夫时间. 这是因为

$$I_{\{S\wedge T>n\}}=I_{\{S>n\}}I_{\{T>n\}}.$$

这样, 如果 T 是马尔可夫时间, 则对任何固定的 $n=0,1,\cdots,T\wedge n=\min\{T,n\}$ 也是.

(c) 如果 S 和 T 是马尔可夫时间, 其较大者 $S\vee T=\max\{S,T\}$ 也是. 因为

$$I_{\{S\vee T\leqslant n\}}=I_{\{S\leqslant n\}}I_{\{T\leqslant n\}}.$$

256

可选抽样定理 [1]

假设 $\{X_n\}$ 是一个鞅, T 是一个关于 $\{Y_n\}$ 的马尔可夫时间. 我们随后将证明

$$E[X_0]=E[X_{T\wedge n}]=\lim_{n\to\infty}E[X_{T\wedge n}].$$

如果 $T<\infty$, 则 $\lim\limits_{n\to\infty}X_{T\wedge n}=X_T$; 实际上, 当 $n>T$ 时, $X_{T\wedge n}=X_T$. 这样, 要是我们证明了当 $n\to\infty$ 时极限与期望可交换, 我们就能导出下面重要的恒等式

$$E[X_0]=\lim_{n\to\infty}E[X_{T\wedge n}]\stackrel{?}{=}E[\lim_{n\to\infty}X_{T\wedge n}]=E[X_T]. \tag{3.1}$$

后面我们要给出几个条件使得这个交换是合理的.

引理 3.1 设 $\{X_n\}$ 是 (上) 鞅, T 是关于 $\{Y_n\}$ 的马尔可夫时间. 则对任意 $n\geqslant k$

$$E[X_nI_{\{T=k\}}](\leqslant)=E[X_kI_{\{T=k\}}]. \tag{3.2}$$

证明 由全概率公式和 (2.9) 得

$$\begin{aligned}E[X_nI_{\{T=k\}}]&=E\{E[X_nI_{\{T=k\}}(Y_0,\cdots,Y_k)|Y_0,\cdots,Y_k]\}\\&=E\{I_{\{T=k\}}E[X_n|Y_0,\cdots,Y_k]\}\\(\leqslant)&=E\{I_{\{T=k\}}X_k\}.\end{aligned}$$

■

1. 也称为**可选停止定理**.

引理 3.2 如果 $\{X_n\}$ 是 (上) 鞅, T 是马尔可夫时间, 则对任意 $n=1,2,\cdots$,

$$E[X_0](\geqslant)=E[X_{T\wedge n}](\geqslant)=E[X_n]. \tag{3.3}$$

证明 利用引理 3.1,

$$\begin{aligned}E[X_{T\wedge n}]&=\sum_{k=0}^{n-1}E[X_TI_{\{T=k\}}]+E[X_nI_{\{T\geqslant n\}}]\\&=\sum_{k=0}^{n-1}E[X_kI_{\{T=k\}}]+E[X_nI_{\{T\geqslant n\}}][\text{当}T=k\text{ 时 }X_T=X_k]\\(\geqslant)&=\sum_{k=0}^{n-1}E[X_nI_{\{T=k\}}]+E[X_nI_{\{T\geqslant n\}}]\quad[\text{根据(3.2)}]\\&=E[X_n].\end{aligned}$$

257

对于鞅, $E[X_n]=E[X_0]$, 在此情况下已证得结论. 对于上鞅情况, 我们证明了 $E[X_{T\wedge n}]\geqslant E[X_n]$, 尚须证明 $E[X_0]\geqslant E[X_{T\wedge n}]$. 我们不妨假设对于所有 n 有 $E[|X_n|]<\infty$. 一般情况可由引理 2.2 后面的注所提示的截断法推得.

如下定义的序列

$$\widetilde{X}_n=\sum_{k=1}^{n}\{X_k-E[X_k|Y_0,\cdots,Y_{k-1}]\}$$

是一个鞅 ($\widetilde{X}_0=0$)[参考 6.1 节例 (b)]. 这样,

$$\begin{aligned}0&=E[\widetilde{X}_{T\wedge n}]\\&=E\Big[\sum_{k=1}^{T\wedge n}\{X_k-E[X_k|Y_0,\cdots,Y_{k-1}]\}\Big]\\&\geqslant E\Big[\sum_{k=1}^{T\wedge n}\{X_k-X_{k-1}\}\Big]=E[X_{T\wedge n}]-E[X_0],\end{aligned}$$

因此,

$$E[X_0]\geqslant E[X_{T\wedge n}],$$

这就完成了证明. ■

现在我们验证式 (3.1) 中极限号与期望可交换. 保证这个交换的可行的最一般条件是随机变量序列 $\{X_{T\wedge n}\}$ 的**一致可积性**, 其意义是

$$\lim_{a\to\infty}\sup_{n\geqslant 0}E[|X^a_{T\wedge n}|]=0,$$

其中

$$X_{T\wedge n}^a = \begin{cases} 0, & \text{如果}|X_{T\wedge n}| < a, \\ X_{T\wedge n}, & \text{如果}|X_{T\wedge n}| \geqslant a, \end{cases}$$

且 $T < \infty$. 下面我们所考虑的虽然不是上述一般条件, 但仍然包含了许多重要的情况.

引理 3.3 设 W 是任意随机变量, 满足 $E[|W|] < \infty$, 并令 T 是马尔可夫时间满足 $\Pr\{T < \infty\} = 1$. 则有

$$\lim_{n\to\infty} E[WI_{\{T>n\}}] = 0 \tag{3.4}$$

和

$$\lim_{n\to\infty} E[WI_{\{T\leqslant n\}}] = E[W]. \tag{3.5}$$

258

证明 我们把这个问题化为非负项级数初等收敛性的问题. 首先

$$\begin{aligned} E[|W|] &\geqslant E[|W|I_{\{T\leqslant n\}}] \\ &= \sum_{k=0}^{n} E[|W||T = k]\Pr\{T = k\} \qquad (\text{全概率公式}) \\ &\xrightarrow{n\to\infty} \sum_{k=0}^{\infty} E[|W||T = k]\Pr\{T = k\} \\ &= E[|W|]. \end{aligned}$$

因此

$$\lim_{n\to\infty} E[|W|I_{\{T\leqslant n\}}] = E[|W|],$$

并且

$$\lim_{n\to\infty} E[|W|I_{\{T>n\}}] = 0.$$

其次, 注意到

$$\begin{aligned} 0 &\leqslant |E[W] - E[WI_{\{T\leqslant n\}}]| \\ &= |E[WI_{\{T>n\}}]| \\ &\leqslant E[|W|I_{\{T>n\}}] \to 0, \end{aligned}$$

这就完成了证明. ■

下面关于受控鞅的可选抽样定理直接由上述引理推得.

定理 3.1 假设 $\{X_n\}$ 是鞅, T 是马尔可夫时间. 如果 $\Pr\{T < \infty\} = 1$ 和 $E\left[\sup_{n\geqslant 0}|X_{T\wedge n}|\right] < \infty$, 则

$$E[X_T] = E[X_0].$$

证明 置 $W = \sup\limits_{n\geqslant 0}|X_{T\wedge n}|$. 由于 $\Pr\{T<\infty\}=1$, 有下面分解式

$$X_T = \sum_{k=0}^{\infty} X_k I_{\{T=k\}} = \sum_{k=0}^{\infty} X_{T\wedge k} I_{\{T=k\}},$$

从而推知 $|X_T| \leqslant W$, 所以 $E[|X_T|] \leqslant E[W] < \infty$, 即 X_T 的期望存在且有限. 我们仅需要证明 $\lim\limits_{n\to\infty} E[X_{T\wedge n}] = E[X_T]$. 我们有

$$\begin{aligned}|E[X_{T\wedge n}] - E[X_T]| &\leqslant E[\ |(X_{T\wedge n} - X_T)|I_{\{T>n\}}] \\ &\leqslant 2E[WI_{\{T>n\}}].\end{aligned}$$

259

但由引理 3.3, 有 $\lim\limits_{n\to\infty} E[WI_{\{T>n\}}] = 0$. 参见式 (3.3) 即知证明已完成. ■

推论 3.1 假设 $\{X_n\}$ 是鞅, T 是关于 $\{Y_n\}$ 的马尔可夫时间, 如果

(i) $E[T] < \infty$,

且存在一个常数 $K<\infty$, 使得对于 $n<T$,

(ii) $E[\ |X_{n+1} - X_n|\ |Y_0,\cdots,Y_n] \leqslant K$, 则 $E[X_T] = E[X_0]$.

证明 定义 $Z_0 = |X_0|$ 和 $Z_n = |X_n - X_{n-1}|, n=1,2,\cdots$, 并令

$$W = Z_0 + \cdots + Z_T,$$

则 $W \geqslant |X_T|$, 且

$$\begin{aligned}E[W] &= \sum_{n=0}^{\infty}\sum_{k=0}^{n} E[Z_k I_{\{T=n\}}] \\ &= \sum_{k=0}^{\infty}\sum_{n=k}^{\infty} E[Z_k I_{\{T=n\}}] \\ &= \sum_{k=0}^{\infty} E[Z_k I_{\{T\geqslant k\}}].\end{aligned}$$

注意到 $I_{\{T\geqslant k\}} = 1 - I_{\{T\leqslant k-1\}}$ 仅是 $\{Y_0,\cdots,Y_{k-1}\}$ 的函数, 并由 (ii), 若 $k\leqslant T$ 则成立不等式 $E[Z_k|Y_0,\cdots,Y_{k-1}] \leqslant K$. 因此,

$$\begin{aligned}\sum_{k=0}^{\infty} E[Z_k I_{\{T\geqslant k\}}] &= \sum_{k=0}^{\infty} E\{E[Z_k I_{\{T\geqslant k\}}|Y_0,\cdots,Y_{k-1}]\} \\ &= \sum_{k=0}^{\infty} E\{I_{\{T\geqslant k\}} E[Z_k|Y_0,\cdots,Y_{k-1}]\} \\ &\leqslant K\sum_{k=0}^{\infty} \Pr\{T\geqslant k\} \\ &\leqslant K(1+E[T]) < \infty.\end{aligned}$$

这样, $E[W]<\infty$. 由 W 的定义知对任意 n 有 $|X_{T\wedge n}|\leqslant W$, 由定理 3.1 即得结论.■ 260

定理 3.2 (可选停止定理). 设 $\{X_n\}$ 是鞅并且 T 是马尔可夫时间. 如果

(i) $\Pr\{T<\infty\}=1$,

(ii) $E[\,|X_T|\,]<\infty$,

(iii) $\lim\limits_{n\to\infty} E[X_n I_{\{T>n\}}]=0$,

则 $E[X_T]=E[X_0]$.

证明 我们强调 (ii) 是必须的, 且由 $E[\,|X_n|\,]<\infty$ 对任意 n 成立不能推得 (ii). 对任意 n, 我们有

$$\begin{aligned}E[X_T]&=E[X_T I_{\{T\leqslant n\}}]+E[X_T I_{\{T>n\}}]\\&=E[X_{T\wedge n}]-E[X_n I_{\{T>n\}}]+E[X_T I_{\{T>n\}}].\end{aligned}$$

由引理 3.2, 有 $E[X_{T\wedge n}]=E[X_0]$, 并且由假设 (iii) 有 $\lim\limits_{n\to\infty}E[X_n I_{\{T>n\}}]=0$. 最后, 利用引理 3.3 于 $W=X_T$, 并由假设 (ii), 推知 $\lim\limits_{n\to\infty}E[X_T I_{\{T>n\}}]=0$. 于是

$$E[X_T]=\lim_{n\to\infty}E[X_{T\wedge n}]=E[X_0],$$

定理证毕. ■

下面是这个基本定理的若干推论.

推论 3.2 假设 $\{X_n\}$ 是鞅, T 是马尔可夫时间, 如果

(i) $\Pr\{T<\infty\}=1$,

并且对于某个 $K<\infty$, 对任意 n,

(ii) $E[X^2_{T\wedge n}]\leqslant K$,

则 $E[X_T]=E[X_0]$.

证明 由于 $X^2_{T\wedge n}\geqslant 0$, 由条件 (ii) 推得

$$\begin{aligned}K&\geqslant E[X^2_{T\wedge n}I_{\{T\leqslant n\}}]\\&=\sum_{k=0}^{n}E[X_T^2|T=k]\Pr\{T=k\}\\&\xrightarrow[n\to\infty]{}\sum_{k=0}^{\infty}E[X_T^2|T=k]\Pr\{T=k\}\\&=E[X_T^2].\end{aligned}$$

261

由施瓦兹不等式推得 $E[\,|X_T|\,]\leqslant(E[X_T^2])^{1/2}<\infty$, 由此验证了定理 3.2 的条件 (ii), 对于 (iii), 我们再次利用施瓦兹不等式得到

$$\begin{aligned}\{E[X_n I_{\{T>n\}}]\}^2&=\{E[X_{T\wedge n}I_{\{T>n\}}]\}^2\\&\leqslant E[X^2_{T\wedge n}]E[I^2_{\{T>n\}}]\\&\leqslant \mathrm{K}\,\Pr\{T>n\}\to 0,\quad 当 n\to\infty.\end{aligned}$$ ■

推论 3.3 假设 $Y_0=0$ 和 $Y_1,Y_2,\cdots$ 是独立同分布随机变量, $E[Y_k]=\mu$ 且 $\mathrm{Var}[Y_k]=\sigma^2<\infty$. 置 $X_n=S_n-n\mu$, 其中 $S_n=Y_0+\cdots+Y_n$. 如果 T 是马尔可夫时间且 $E[T]<\infty$, 则 $E[\,|X_T|\,]<\infty$ 并且 $E[X_T]=E[S_T]-\mu E[T]=0$.

证明 我们应用定理于鞅 $\{S_n-n\mu\}$. 令 $Y_0'=0, Y_k'=Y_k-\mu, k=1,2,\cdots$. 为证明 $E[\,|X_T|\,]<\infty$, 我们有

$$\begin{aligned}E[\,|X_T|\,]&\leqslant E\Big[\sum_{k=1}^{T}|Y_k'|\Big]\\&=E\Big[\sum_{n=1}^{\infty}\sum_{k=1}^{n}|Y_k'|I_{\{T=n\}}\Big]\\&=E\Big[\sum_{k=1}^{\infty}|Y_k'|\,I_{\{T\geqslant k\}}\Big].\end{aligned}$$

因为 $I_{\{T\geqslant k\}}=I_{\{T>k-1\}}$ 仅仅依赖于 $\{Y_0,\cdots,Y_{k-1}\}$, 于是独立于 Y_k'. 因此

$$\begin{aligned}E\Big[\sum_{k=1}^{\infty}|Y_k'|\,I_{\{T\geqslant k\}}\Big]&=E[\,|Y_1'|\,]\sum_{k=1}^{\infty}\Pr\{T\geqslant k\}\\&=E[\,|Y_1'|\,]E[T]<\infty.\end{aligned}$$

为验证定理 3.2 的条件 (iii), 利用施瓦兹不等式有,

$$\begin{aligned}(E[X_nI_{\{T>n\}}])^2&\leqslant E[X_n^2]E[I_{\{T>n\}}]\\&\leqslant n\sigma^2\Pr\{T\geqslant n\}.\end{aligned}$$

但, $\infty>E[T]=\sum_{k=0}^{\infty}k\Pr\{T=k\}$, 所以

$$\begin{aligned}0&=\lim_{n\to\infty}\sum_{k\geqslant n}k\Pr\{T=k\}\\&\geqslant\lim_{n\to\infty}n\Pr\{T\geqslant n\}\geqslant 0.\end{aligned}$$

262

■

6.4 可选抽样定理的若干应用

我们将看到, 可选抽样定理很快可以在与随机过程有关的一些概率的计算和估计问题中找到应用. 涉及布朗运动的更多应用, 将在 7.5 节介绍.

(a) 随机游动

由可选抽样定理可迅速推得与随机游动有关的许多重要结果. 首先我们讨论本章开头的一个例子. 令 $Y_0=0$, 对 $i=1,2,\cdots$, 令 Y_i 是独立同分布随机变量, 且 $\Pr\{Y_i=1\}=p$, $\Pr\{Y_i=-1\}=q=1-p$. 令 $S_0=0, S_n=Y_1+\cdots+Y_n, n\geqslant 1$.

先假设 $p=q=\frac{1}{2}$. 则 $\{S_n\}$ 是一个鞅. 如果 $T=\min\{n:S_n=1\}$, 则 $\Pr\{T<\infty\}=1$, 因为 S_n 是常返的. 但 $S_T\equiv 1$, 所以 $E[S_T]\neq E[S_0]=0$. 这就与推论 3.3 的结果矛盾. 这样, 此推论的假设不能成立, 特别, 我们有 $E[T]=\infty$.

继续假设 $p=q=\frac{1}{2}$, 但现在令

$$T=\min\{n:S_n=-a \text{ 或 } S_n=b\},\quad (a,b\text{为正整数}).$$

令 v_a 是 S_n 达到 b 之前到达 $-a$ 的概率, 则由定理 3.1 有

$$0=E[S_T]=v_a(-a)+(1-v_a)b$$

或

$$v_a=\frac{b}{a+b},$$

这个结果在第 3 章是由其他方法得到的.

$Z_n=S_n^2-n$ 也是一个鞅, 并且

$$E[Z_T]=0=[v_a a^2+(1-v_a)b^2]-E[T],$$

由此得到

$$E[T]=ab.$$

现在假设 $p>q$, 并置 $\mu=E[Y_k]=p-q>0$. 则

$$X_n=S_n-n\mu \tag{4.1}$$

和

$$X_n'=(q/p)^{S_n} \tag{4.2}$$

263

是鞅. 由 (4.1) 和推论 3.1 我们得到, 对任何马尔可夫时间 T, 若 $E[T]<\infty$, 有

$$E[S_T]=\mu E[T].$$

为利用 (4.2), 令

$$T=\min\{n:S_n=-a \text{ 或 } S_n=b\}.$$

故

$$1=E[X_T']=v_a\left(\frac{q}{p}\right)^{-a}+(1-v_a)\left(\frac{q}{p}\right)^{b},$$

或

$$v_a=\frac{1-(q/p)^b}{(q/p)^{-a}-(q/p)^b},$$

同前面一样, 其中 v_a 是 S_n 到达 b 之前到达 $-a$ 的概率. 这个结果又和我们在第 3 章用其他方法导出的公式是一致的.

(b) Wald 恒等式

设 $Y_0=0$ 和 $Y_1,Y_2,\cdots$ 是非退化独立同分布随机变量, 并且**矩母函数**.

$$\phi(\theta)=E[\exp\{\theta Y_1\}]$$

对于 θ 在某个包含原点的开区间内有定义且有限. 置 $S_0=0$ 和 $S_n=Y_1+\cdots+Y_n$. 令 $-a<0$ 和 $b>0$, 并设

$$T=\min\{n:S_n\leqslant -a \text{ 或 } S_n\geqslant b\}.$$

Wald 基本恒等式是

$$E[\phi(\theta)^{-T}\exp\{\theta S_T\}]=1, \tag{4.3}$$

对于满足 $\phi(\theta)\geqslant 1$ 的任意 θ 成立. 这个恒等式在应用概率和统计中有大量的应用.

我们将利用推论 3.1 证实 (4.3). 由 6.1 节例 (g) 知道 $X_0=1$ 并且

$$X_n=\phi(\theta)^{-n}\exp\{\theta S_n\}, n\geqslant 1$$

定义了一个关于 $\{Y_n\}$ 的鞅, 则

$$\begin{aligned}&E[\,|X_{n+1}-X_n|\mid Y_0,\cdots,Y_n]\\&=X_nE[\,|\phi(\theta)^{-1}\exp(\theta Y_{n+1})-1|\,].\end{aligned}$$

264

由假设, $\phi(\theta)^{-n}\leqslant 1$, 并且对于 $n<T, \exp\{\theta S_n\}\leqslant \mathrm{e}^b$. 这样, $X_n\leqslant \mathrm{e}^b$ 对于 $n<T$ 成立. 此外,

$$\begin{aligned}&E[\,|\phi(\theta)^{-1}\exp(\theta Y_{n+1})-1|\,]\\&\leqslant \phi(\theta)^{-1}\{E[\exp(\theta Y_{n+1})]+\phi(\theta)\}=2.\end{aligned}$$

于是, 对于 $n<T$,

$$E[\,|X_{n+1}-X_n|\,|Y_0,\cdots,Y_n\,]\leqslant 2\mathrm{e}^b,$$

为了应用推论 3.1, 我们只需验证 $E[T]<\infty$. 令 $c=a+b$. 由于 Y_k 假定是非退化的, 则存在一个整数 N 和 $\delta>0$ 使得 $\Pr\{\,|S_N|>c\,\}>\delta$. 令 $S_1'=S_N, S_2'=S_{2N}-S_N,\cdots,S_k'=S_{kN}-S_{(k-1)N}$. 则

$$\begin{aligned}\Pr\{T\geqslant kN\}&\leqslant \Pr\{\,|S_1'|\leqslant c\,\}\cdots\Pr\{\,|S_k'|\leqslant c\,\}\\&\leqslant(1-\delta)^k,\end{aligned}$$

因 $\Pr\{T \geqslant n\}$ 关于 n 是递减的, 我们得到

$$E[T] = \sum_{n=1}^{\infty} \Pr\{T \geqslant n\} \leqslant N \sum_{k=0}^{\infty} \Pr\{T \geqslant kN\} \leqslant N/\delta < \infty.$$

为了说明 Wald 等式是被普遍应用的, 我们假设存在一个值 $\theta_0 \neq 0$ 使 $\phi(\theta_0) = 1$. 此时 (4.3) 变为

$$E[\exp\{\theta_0 S_T\}] = 1.$$

置

$$E_a = E[\exp\{\theta_0 S_T\}|S_T \leqslant -a]$$

和

$$E_b = E[\exp\{\theta_0 S_T\}|S_T \geqslant b],$$

我们得到

$$\begin{aligned} 1 &= E_a \Pr\{S_T \leqslant -a\} + E_b \Pr\{S_T \geqslant b\} \\ &= E_a + (E_b - E_a)\Pr\{S_T \geqslant b\}, \end{aligned}$$

或

$$\Pr\{S_T \geqslant b\} = \frac{1 - E_a}{E_b - E_a}.$$

倘若 S_n 离开区间 $[-a, b]$, 我们可以期望 $E_a \cong \exp\{-\theta_0 a\}$ 和 $E_b \cong \exp\{\theta_0 b\}$. 当离开边界不是太远的情况下, 情况确实如此, 这样我们就直观地得到 Wald 的近似式

$$\Pr\{S_T \geqslant b\} \cong \frac{1 - \exp\{-\theta_0 a\}}{\exp\{\theta_0 b\} - \exp\{-\theta_0 a\}}.$$

265

现在我们回到 (4.3) 并关于 θ 形式地微分它, 得到

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}\theta} E[\phi(\theta)^{-T} \exp\{\theta S_T\}] \\ &= E[(-T\phi(\theta)^{-T-1}\phi'(\theta) + \phi(\theta)^{-T} S_T)\exp\{\theta S_T\}] \\ &= -\phi'(\theta) E[T\phi(\theta)^{-T-1}\exp\{\theta S_T\}] + E[\phi(\theta)^{-T} S_T \exp\{\theta S_T\}]. \end{aligned}$$

置 $\theta = 0$, 利用 $\phi(0) = 1$ 和 $\phi'(0) = E[Y_1]$, 则

$$0 = -E[Y_1]E[T] + E[S_T]$$

或

$$E[S_T] = E[Y_1]E[T].$$

关于上鞅的可选停止定理

设 $\{X_n\}$ 是关于 $\{Y_n\}$ 的上鞅. 依照引理 3.2, $E[X_0] \geqslant E[X_{T\wedge n}]$ 对于任何马尔可夫时间成立. 与鞅的情况类似倘若当 $n \to \infty$ 时, 我们能够交换极限与期望, 我们即可推得 $E[X_0] \geqslant E[X_T]$. 下面的两个定理是这方面的两个重要情形.

定理 4.1 设 $\{X_n\}$ 是上鞅, T 是马尔可夫时间. 如果 $\Pr\{T < \infty\} = 1$ 并且存在一个随机变量 $W \geqslant 0$ 使得 $E[W] < \infty$ 和 $X_{T\wedge n} > -W$ 对任意 n 成立, 则

$$E[X_0] \geqslant E[X_T].$$

证明 设 $c > 0$ 是固定的, 并定义 $X_n^c = \min\{c, X_n\}, n = 0, 1, 2, \cdots$, 则 $\{X_n^c\}$ 关于 $\{Y_n\}$ 也是上鞅 (参见 6.2 节), 所以

$$E[X_0^c] \geqslant E[X_{T\wedge n}^c],$$

由于

$$|X_{T\wedge n}^c| \leqslant \max\{c, W\}$$

对任意 n 成立, 与定理 3.1 一样当 $n \to \infty$ 时, 我们可以交换极限与期望, 导出

$$E[X_0^c] \geqslant E[X_T^c].$$

然而, 显然 $E[X_0] \geqslant E[X_0^c]$, 而

266

$$\lim_{c\to\infty} E[X_T^c] = \lim_{c\to\infty} \int_{-\infty}^{c} x \mathrm{d}\Pr\{X_T \leqslant x\} = E[X_T].$$

这样, 即得到 $E[X_0] \geqslant E[X_T]$. ■

定理 4.2 设 $\{X_n\}$ 是上鞅, T 是马尔可夫时间. 如果对所有 n 有 $X_n \geqslant 0$, 则

$$E[X_0] \geqslant E[X_T I_{\{T<\infty\}}].$$

证明 照例, 有

$$E[X_0] \geqslant E[X_{T\wedge n}].$$

由于 $X_n \geqslant 0, X_{T\wedge n} = X_T I_{\{T\leqslant n\}} + X_n I_{\{T>n\}} \geqslant X_T I_{\{T\leqslant n\}}$, 所以 $E[X_0] \geqslant E[X_T I_{\{T\leqslant n\}}]$, 并且

$$\begin{aligned} E[X_0] &\geqslant \lim_{n\to\infty} E[X_T I_{\{T\leqslant n\}}] \\ &= \lim_{n\to\infty} \sum_{k=0}^{n} E[X_T | T = k] \Pr\{T = k\} \\ &= \sum_{k=0}^{\infty} E[X_T | T = k] \Pr\{T = k\} \\ &= E[X_T I_{\{T<\infty\}}]. \end{aligned}$$

■

权证价值问题

设 W_n 为一种资产在第 n 天在一个公共市场上交易的价格，比如股市上的股票. 设 Y_n 为在第 n 天的价格与第 $n-1$ 天的价格的比率，因此，$W_n = w \times Y_1 \times \cdots \times Y_n$，其中 $W_0 = w$ 为当天的价格. 在这种情况下，根据历史上著名且富有争议的"随机游动假设"，可以断定 $Y_1, Y_2, \cdots$ 是独立同分布的正随机变量. 可以证明，这种刻画"理想市场"的假设导致了这样的结果. 目前，人们感兴趣的是用条件比较弱的鞅理论代替随机游动假设. 设长期增长的通货膨胀率 $\alpha \geqslant 0$ 以及 $\mathrm{e}^{-\alpha n}W_n$ 为关于 Y_n 的鞅. 因此，$E[\mathrm{e}^{-\alpha n}W_n] = E[W_0] = w$ 或者 $E[W_n] = w\mathrm{e}^{\alpha n}$，所以 α 表示资产市场价格的平均增长率.

现在考虑一个权证合约，即在任何时刻权证持有者都可以购买资产，只要资产的价格在他乐意接收的范围内，而不去考虑市场的价格为多少. 在股票市场，这种选择权被称为"权证". 如果必要，通过改变价值的单位，我们假设约定的价格为 1. 因此，如果 $W_n > 1$，则权证的所有者就可以行权，以约定的价格购买资产，然后以价格 W_n 在市场上卖出，以赚取利润 $W_n - 1$. 如果 $W_n \leqslant 1$，就没有利润可言了. 因此，权证持有者潜在的利润可以表示为 [267]

$$r(W_n) = (W_n - 1)^+ = \max\{W_n - 1, 0\}.$$

另一种保存选择权是保留资产，因此平均的回报率为 α. 这种选择通常是有用的，当回报率 $\beta > \alpha$ 时，人们通常采用这种选择. 等价地，第 n 天的潜在回报 $r(W_n)$ 由因子 $\mathrm{e}^{-\beta n}$ 决定. 设 T 为行权时间. 在这个一般鞅模型中，可以得到平均贴现利润 $E[\mathrm{e}^{-\beta T}r(W_T)]$ 的界限，它是关于目前价格 w 的函数，在矩条件下存在 $\theta > 1$ 使得

$$E[Y_n^\theta | Y_1, \cdots, Y_{n-1}] \leqslant \mathrm{e}^\beta, \qquad n = 1, 2, \cdots \tag{4.4}$$

实际上，(4.4) 式是唯一需要的假设. 对于市场价格我们的结果既不依赖于随机游动模型也不依赖于鞅模型，而是这两个模型的结合. 注意到，当 $T = \infty$ 时，约定 $\mathrm{e}^{-\beta T}r(W_T) = 0$，因此如果都不行权将无利可图.

例 设 $Y_1, Y_2, \cdots$ 为独立同分布的对数正态随机变量，也就是 $V_k = \ln Y_k$ 服从均值为 μ，方差为 σ^2 的正态分布，则

$$E[Y_n | Y_1, \cdots, Y_{n-1}] = E[Y_n] = E[\exp\{V_n\}] = \exp\left\{\mu + \frac{1}{2}\sigma^2\right\},$$

因此 $\alpha = \mu + \dfrac{1}{2}\sigma^2$，而

$$E[Y_n^\theta | Y_1, \cdots, Y_{n-1}] = E[\exp\{\theta V_n\}] = \exp\left\{\theta\mu + \frac{1}{2}\theta^2\sigma^2\right\}.$$

我们解关系式 $\theta\mu+\frac{1}{2}\theta^2\sigma^2=\beta$, 可得 (4.4) 式等号成立的条件是

$$\theta=\frac{(\mu^2+2\sigma^2\beta)^{1/2}-\mu}{\sigma^2}$$

因为 $\beta>\alpha=\mu+\frac{1}{2}\sigma^2$,

$$\theta>\frac{(\mu^2+2\sigma^2\alpha)^{1/2}-\mu}{\sigma^2}=1,$$

正如我们所需.

回到一般情况, 我们将证明对所有 $w\geqslant 0, E[\mathrm{e}^{-\beta T}r(W_T)]\leqslant f(w)$, 其中

$$f(w)=\begin{cases} w^\theta(\theta-1)^{\theta-1}/\theta^\theta, & \text{若 } w\leqslant\theta/(\theta-1),\\ w-1, & \text{若 } w>\theta/(\theta-1).\end{cases}$$

268

注意到, 对所有 $w>0$, 有 $f(w)\leqslant w^\theta(\theta-1)^{\theta-1}/\theta^\theta$. 这个界限并不会影响权证所有者何时做出何种策略. 从而这个界限可能被出售资产的卖方用来降低他们的平均损失. 同时注意到, 如果当天的价格 $W_0=w$ 超过 $\theta/(\theta-1)$, 则权证的价值至多为 $f(w)=w-1$, 这个数量只是在瞬间到达的, 因此当市场的价格到达这个高度时就应该行权. 这个结论只要求满足 (4.4) 式的矩假设, 而不需要考虑每天价格改变形式的概率分布.

为了证明这个界限, 设 $X_n=\mathrm{e}^{-\beta n}f(W_n)$, 我们断言 $\{X_n\}$ 是一个关于 $\{Y_n\}$ 的非负上鞅. 只需证明对所有$w\geqslant 0$,

$$f(w)\geqslant \mathrm{e}^{-\beta}E[f(w\times Y)], \tag{4.5}$$

$$E[Y^\theta]\leqslant \mathrm{e}^\beta, \tag{4.6}$$

只要因为只要证明了 (4.5) 式, 使用 (4.4) 式, 我们就可以得到

$$\begin{aligned} E[X_n|Y_0,\cdots,Y_{n-1}]&=\mathrm{e}^{-\beta n}E[f(W_{n-1}\times Y_n)|Y_0,\cdots,Y_{n-1}]\\ &\leqslant \mathrm{e}^{-\beta(n-1)}f(W_{n-1})=X_{n-1}.\end{aligned}$$

下面来证明 (4.5). 对固定的 $t>1$, 定义

$$v(t,w)=w^t(t-1)^{t-1}/t^t,\qquad w\geqslant 0.$$

则

$$v(t,w)\geqslant f(w),\qquad 1<t\leqslant\theta, w\geqslant 0. \tag{4.7}$$

我们将 (4.7) 的证明留给读者完成. [在$g_t(w)=v(t,w)-(w-1)$和$w_0=t/(t-1)>\theta/(\theta-1)$ 条件下, 读者首先需要证明 $g_t(w_0)=0$ 和 $g_t(w_0)=0$, 这里, 表示对 w 的

导数. 然后, 证明二阶导数 $g_t''(w_0) > 0$, 读者就知道对所有 $w, g_t(w) \geqslant 0$ 或 $v(t,w) \geqslant (w-1)$. 其次, 读者将 $\ln v(t,w)$ 对 t 求导并化简得

$$\frac{\mathrm{d}\ln v(t,w)}{\mathrm{d}t} = \ln\left\{\frac{w}{t/(t-1)}\right\}$$

当 $w < \theta/(\theta-1) < t/(t-1)$ 时上式为负的, 也就是对于 $w \leqslant \theta/(\theta-1), v(t,w)$ 随着 t 从 θ 递减而增加. 因此, 当 w 在这个范围内时 $v(t,w) > v(\theta,w) = f(w)$.]

269

接着, 我们考虑两种情况. 首先假设 $w \leqslant \theta/(\theta-1)$, 则

$$\begin{aligned} f(w) &= v(\theta,w) \\ &\geqslant \mathrm{e}^{-\beta} v(\theta,w) E[Y^\theta] && [\text{由}(4.6)] \\ &= \mathrm{e}^{-\beta} E[v(\theta, w\times Y)] && [\text{由}v(\theta,w)\text{的定义}] \\ &\geqslant \mathrm{e}^{-\beta} E[f(w\times Y)] && [\text{由}(4.7)]. \end{aligned} \tag{*}$$

第二种情况是 $w > \theta/(\theta-1)$. 现在, $E[Y^\alpha]$ 是关于 $\alpha \in [0,\theta]$ 的凸函数, 其中 $E[Y^0] = 1 < \mathrm{e}^\beta$. 利用 Jensen 不等式和 (4.6) 的假设, 对于 $w/(w-1) < \theta$ 有 $E[Y^{w/(w-1)}] \leqslant \mathrm{e}^\beta$. 现在考虑

$$\begin{aligned} f(w) &= w-1 \\ &= v(w/(w-1), w) \\ &\geqslant \mathrm{e}^{-\beta} E[v(w/w-1, w\times Y)] && [\text{因为}E(Y^{w/w-1}) \leqslant \mathrm{e}^\beta] \\ &\geqslant \mathrm{e}^{-\beta} E[f(w\times Y)] && [\text{由}(4.7)]. \end{aligned} \tag{**}$$

通过 (*) 和 (**) 可以证明 (4.5). 因此, $X_n = \mathrm{e}^{-\beta n} f(W_n)$ 构成一个非负上鞅, 并且对任意的马尔可夫时间 T.

$$X_0 = f(w) \geqslant E[\mathrm{e}^{-\beta T} f(W_T)], \tag{4.8}$$

最后, 我们验证 $f(w) \geqslant r(w) = (w-1)^+$, 由 (4.8) 式可推得

$$f(w) \geqslant E[\mathrm{e}^{-\beta T} r(W_T)],$$

正如前面所断言的. 很明显地, 对于所有 $w, f(w) \geqslant 0$, 以及对于 $w \geqslant \theta/(\theta-1), f(w) = w-1$. 进一步地,

$$\mathrm{d}f/\mathrm{d}w = [(\theta-1)/\theta]^{\theta-1} w^{\theta-1} < 1, \qquad \text{对于} w < \theta/(\theta-1).$$

通过积分可以得到

$$f\left(\frac{\theta}{\theta-1}\right) - f(w) < \frac{\theta}{\theta-1} - w,$$

因此, 对于 $w < \theta/(\theta-1)$, 有 $f(w) > w-1$, 这就完成了对不等式 $f(w) \geqslant (w-1)^+$ 的证明.

水库问题

令 Z_t 表示容量为 b 的水坝在时刻 t 的贮水量, I_t 为时间区间 $(t,t+1]$ 内随机流入量, O_t 为其流出量. 由于蓄水量不能是负的, 也不能超过容量 b, 因此有平衡方程

270

$$Z_{t+1} = \min\{(Z_t + I_t - O_t)^+, b\}$$

此处 "x^+" 表示 "$\max\{x,0\}$".

假设由于航行、游览或紧急情况的需要, 必须保持水库的蓄水量高于某最低可接受的水平 a. 从而

$$T = \min\{t : Z_t \leqslant a\}$$

是不满足要求的首达时刻. 需要量 $\{O_t\}$ 和容量 b 是可控制参数或设计参数, 对于它们不同的值, 水坝性能可用平均时间 $E[T]$ 来比较, $E[T]$ 越大则水坝性能越好.

通常办法是对流入量和流出量做较准确的假设, 然后计算 $E[T]$. 然而, 对于水库系统验证这些具体的假设通常是困难的, 由于季节的影响和上游地表水的贮存, 可在一段时间内影响水库的流入量. 并且关于流入量的分布, 可供使用的信息一般很少, 虽然有过去资料, 但它与现在的相关性很小, 因为在整个流域系统地形的改变是经常发生的.

在这个例子中, 我们将在比较弱的一般假设下得到平均时间 $E[T]$ 的界. 设在已知过去的条件下, 净流入量

$$Y_{t+1} = I_{t+1} - O_{t+1}$$

的条件分布满足

$$E[Y_{t+1}|Y_1,\cdots,Y_t] \leqslant m \tag{4.9}$$

和

$$E[\exp\{-\lambda Y_{t+1}\}|Y_1,\cdots,Y_t] \leqslant 1, \tag{4.10}$$

其中 m 和 λ 为已知正常数. 下面我们来证明 $E[T] \geqslant f(z)$, 其中

$$f(z) = \frac{1}{m}\left\{e^{\lambda(b-a)}\frac{1-e^{-\lambda(z-a)}}{\lambda} - (z-a)\right\}, a \leqslant z \leqslant b,$$

且 $Z_0 = z$ 是初始水坝蓄水量. 保守的设计者把 $f(z)$ 作为其设计标准, 因为它表示最坏的情况, 即蓄水量达到临界水平 a 的平均时间最短.

我们断言

271

$$X_n = f(Z_n) + n$$

构成一个关于 $\{Y_n\}$ 的下鞅. 延拓 f 的定义域, 对于 $z<a$, 令 $f(z)=0$, 对于 $z>b$, 令 $f(z)=f(b)$. 因而 $f(Z_{n+1})=f(Z_n+Y_{n+1})$. 对任意 z, 置

$$g(z)=\frac{1}{m}\left\{\mathrm{e}^{\lambda(b-a)}\frac{1-\mathrm{e}^{-\lambda(z-a)}}{\lambda}-(z-a)\right\}.$$

由于 $g(z)$ 一直递增至 $z=b$, 然后递减, 因此对所有 z 我们有 $f(z)\geqslant g(z)$, 并且

$$\begin{aligned}E[X_{n+1}|Y_0,\cdots,Y_n]&=E[f(Z_{n+1})|Y_0,\cdots,Y_n]+(n+1)\\&=E[f(Z_n+Y_{n+1})|Y_0,\cdots,Y_n]+(n+1)\\&\geqslant E[g(Z_n+Y_{n+1})|Y_0,\cdots,Y_n]+(n+1).\end{aligned}$$

如果 U 是随机变量且满足 $E[U]\leqslant m$ 和 $E[\mathrm{e}^{-\lambda U}]\leqslant 1$, 则对于 $a\leqslant z\leqslant b$ 成立

$$\begin{aligned}E[g(z+U)]&=\frac{1}{m}\left\{\mathrm{e}^{\lambda(b-a)}\frac{1-E[\mathrm{e}^{-\lambda(z+U-a)}]}{\lambda}-(z+E[U]-a)\right\}\\&\geqslant f(z)-1,\end{aligned}$$

利用以上的分析和假设 (4.9)、(4.10), 我们推得

$$E[X_{n+1}|Y_0,\cdots,Y_n]\geqslant f(Z_n)+n=X_n,$$

即 X_n 是下鞅. 令 $T=\min\{n:Z_n\leqslant a\}$ 是蓄水量首达临界高度 a 的时刻. 由下鞅的可选抽样定理, 有

$$E[f(Z_{T\wedge n})+(T\wedge n)]\geqslant f(z),$$

其中 $z=Z_0$ 是初始水坝蓄水量. 由于 f 是有界函数, 利用引理 3.3 易得

$$\lim_{n\to\infty}E[f(Z_{T\wedge n})]=E[f(Z_T)]=0,\quad (\text{由于}Z_T\leqslant a)$$

和

$$\begin{aligned}\lim_{n\to\infty}E[T\wedge n]&=\lim_{n\to\infty}\sum_{k=1}^{n}\Pr\{T\geqslant k\}\\&=\sum_{k=1}^{\infty}\Pr\{T\geqslant k\}=E[T].\end{aligned}$$

因此

$$\begin{aligned}f(z)&\leqslant\lim_{n\to\infty}\{E[f(Z_{T\wedge n})]+E[T\wedge n]\}\\&=E[T],\end{aligned}$$

这就是要证明的. 272

若 $Y_1,Y_2,\cdots,Y_i,\cdots$ 是均值为 $\mu_i\leqslant m$, 方差 $\sigma_i^2\leqslant 2\mu_i/\lambda$ 的独立正态随机变量, 则条件 (4.9)、(4.10) 必定满足.

上穿不等式

给定一个关于 $\{Y_n\}$ 的下鞅 $\{X_n\}$, 以及实数 $a < b$, 正整数 N, 定义 $V_{a,b}$ 是满足下列条件的数对 (i,j) 的数目: $0 \leqslant i < j \leqslant N, X_i \leqslant a, a < X_k < b, i < k < j$ 且 $X_j \geqslant b$. 因此, $V_{a,b}$ 表示 X_n 对于 $n = 0, 1, \cdots, N$ 上穿区间 (a,b) 的次数, 即从低于水平线 a 上穿到水平线 b 之上的次数. 我们将证明重要的上穿不等式

$$E[V_{a,b}] \leqslant \frac{E[(X_N - a)^+] - E[(X_0 - a)^+]}{b-a}. \tag{4.11}$$

上穿不等式指出下鞅所可能有的振动的限度, 并且提示了样本轨道的粗略性态. 这个不等式及其推广和变形在概率分析中被广泛用于证明收敛性定理和研究连续参数随机过程的样本轨道的增长和连续性.

我们需要引理 3.2 如下形式的推广, 在下鞅情形它同时包含了两个马尔可夫时间. 这是更加一般的可选抽样定理的特别情况, 更一般的可选抽样定理是包含几个甚至可数多个马尔可夫时间的.

引理 4.1 设 $\{X_n\}$ 是下鞅, S, T 是关于 $\{Y_n\}$ 的马尔可夫时间. 假设 $0 \leqslant S \leqslant T \leqslant N$, 其中 N 是固定正整数. 则

$$E[X_S] \leqslant E[X_T].$$

证明 令 k 是固定的, 对于 $k \leqslant n \leqslant N$, 因为 $I_{\{T>n\}}$ 仅依赖于 $Y_0, \cdots, Y_n$, 利用条件概率的性质, 我们有

$$\begin{aligned}
&E[X_{n+1}I_{\{T>n\}}I_{\{S=k\}}] \\
&\quad = E[E\{X_{n+1}|Y_0, \cdots, Y_n\}I_{\{T>n\}}I_{\{S=k\}}] \\
&\quad \geqslant E[X_n I_{\{T>n\}}I_{\{S=k\}}].
\end{aligned}$$

这样,

$$\begin{aligned}
E[X_{T\wedge n}I_{\{S=k\}}] &= E[X_T I_{\{T\leqslant n\}}I_{\{S=k\}}] + E[X_n I_{\{T>n\}}I_{\{S=k\}}] \\
&\leqslant E[X_T I_{\{T\leqslant n\}}I_{\{S=k\}}] + E[X_{n+1}I_{\{T>n\}}I_{\{S=k\}}] \\
&= E[X_{T\wedge(n+1)}I_{\{S=k\}}],
\end{aligned}$$

273

这是 n 的单调增函数. 置 $n = k$, 然后 $n = N$, 利用假设 $S \leqslant T \leqslant N$ 导出关系式

$$E[X_k I_{\{S=k\}}] \leqslant E[X_T I_{\{S=k\}}].$$

因此,

$$\begin{aligned}
E[X_S] &= \sum_{k=0}^{N} E[X_S I_{\{S=k\}}] \\
&= \sum_{k=0}^{N} E[X_k I_{\{S=k\}}] \\
&\leqslant \sum_{k=0}^{N} E[X_T I_{\{S=k\}}] \\
&= E[X_T],
\end{aligned}$$

故引理证毕. ■

我们应用引理来证明上穿不等式, 定义

$$\widehat{X}_n = (X_n - a)^+.$$

由于 $g(x) = (x-a)^+ = \max\{(x-a), 0\}$ 是 x 的凸的增函数, 引理 2.2 告诉我们 $\{\widehat{X}_n\}$ 也是关于 $\{Y_n\}$ 的下鞅. 定义 $T_1 \equiv 0$, 且对于 $k = 1, 2, \cdots, N$, 若 k 是偶数, 设

$$T_k = \begin{cases} N, & \text{如果}\widehat{X}_j \neq 0,\ \text{对}j > T_{k-1}, \\ \min\{j : j > T_{k-1}, \widehat{X}_j = 0\}, & \text{其他;} \end{cases}$$

若 k 是奇数, 设

$$T_k = \begin{cases} N, & \text{如果}\widehat{X}_j < b - a \text{ 对所有 } j > T_{k-1}, \\ \min\{j : j > T_{k-1}, \widehat{X}_j \geqslant b - a\}, & \text{其他.} \end{cases}$$

又置 $T_{N+1} = N$, 则每个 T_k 是马尔可夫时间 (请严格证明之), 并且 $T_k \leqslant T_{k+1}$, 于是, 利用刚刚得到的引理可知

$$E[\widehat{X}_{T_k}] \leqslant E[\widehat{X}_{T_{k+1}}]. \tag{4.12}$$

因此

$$\begin{aligned}
\widehat{X}_N - \widehat{X}_0 &= \sum_{k=1}^{N} \left(\widehat{X}_{T_{k+1}} - \widehat{X}_{T_k}\right) \\
&= \sum_{k=2,4,\cdots} \left(\widehat{X}_{T_{k+1}} - \widehat{X}_{T_k}\right) + \sum_{k=1,3,\cdots} \left(\widehat{X}_{T_{k+1}} - \widehat{X}_{T_k}\right).
\end{aligned}$$

274

现在, 如果 k 是偶数, 只要出现一次上穿则 $\widehat{X}_{T_{k+1}} - \widehat{X}_{T_k}$ 不等于零, 因此至少是 $(b-a)$, 且由 (4.12) 知第二个和的期望值是非负的. 这样一来,

$$E[\widehat{X}_N - \widehat{X}_0] \geqslant (b-a)E[V_{a,b}],$$

或

$$E[V_{a,b}] \leqslant \frac{E[(X_N - a)^+] - E[(X_0 - a)^+]}{b - a},$$

证毕. ■

部分和不等式

设随机变量 $X_1, X_2, \cdots$ 具有有限条件矩

$$M_k = E[X_k | X_1, \cdots, X_{k-1}]$$

和

$$V_k = E[(X_k - M_k)^2 | X_1, \cdots, X_{k-1}].$$

令 $S_n = X_1 + \cdots + X_n (S_0 = 0)$. 若对所有 k 满足条件

$$M_k \leqslant -\alpha V_k, \tag{4.13}$$

其中 α 为某个固定正数, 我们将证明当 $x < l$ 时成立

$$\Pr\left\{\sup_{n \geqslant 0}[x + S_n] > l\right\} \leqslant \frac{1}{1 + \alpha(l - x)}. \tag{4.14}$$

假如被加项为具有均值 $\mu < 0$ 及方差 σ^2 的独立同分布随机变量, 由大数定律, 当 $n \to \infty$ 时有 $S_n \to -\infty$ 并且 $M = \max_{n \geqslant 0} S_n < \infty$(见第 17 章). 在不等式 (4.14) 中可取 $\alpha = |\mu|/\sigma^2$, 从而得到

$$\Pr\{M > l\} \leqslant \sigma^2/[\sigma^2 + |\mu| l].$$

现在回到一般情况, 我们不假设被加项是独立的. 定义

$$f(z) = \begin{cases} \dfrac{1}{1 + \alpha(l - z)}, & \text{当 } z < l, \\ 1, & \text{当 } z \geqslant l. \end{cases}$$

275

我们的目的是在条件 (4.13) 之下证明 $\{f(x + S_n)\}$ 为一非负上鞅. 若证得此事实, 令 T 是使得 $x + S_n > l$ 成立的第一个 n 值, 应用定理 4.2, 即有

$$\begin{aligned} f(x) &\geqslant E[f(x + S_T) I_{\{T < \infty\}}] \\ &= \Pr\left\{\sup_{n \geqslant 0}[x + S_n] > l\right\}. \end{aligned}$$

为后面证明方便, 记 $f(y)$ 的导数表达式如下:

$$f'(y) = \frac{\alpha}{[1 + \alpha(l - y)]^2} = \alpha[f(y)]^2, \qquad \text{当 } y < l, \tag{4.15}$$

$$f''(y) = 2\alpha f(y)f'(y), \qquad \text{当 } y < l, \tag{4.16}$$

由于 $0 \leqslant f(y) < 1$, 从而

$$f''(y) < 2\alpha f'(y), \qquad \text{当} y < l. \tag{4.17}$$

固定一任意点 $z < l$. 令 $g(y)$ 是 y 的二次式且与 $f(y)$ 切于 $y = z$. 因而 $g(y)$ 具有形式

$$g(y) = f(z) + f'(z)(y - z) + a(y - z)^2,$$

其中 a 是一适当常数, 现我们要求 $g(l) = f(l) = 1$, 为此只需取

$$a = \alpha f'(z).$$

事实上,

$$\begin{aligned} g(l) &= f(z) + f'(z)(l - z) + \alpha f'(z)(l - z)^2 \\ &= f(z)\{1 + \alpha(l - z)[1 + \alpha(l - z)]f(z)\} \quad (\text{由}(4.15)) \\ &= f(z)\{1 + \alpha(l - z)\} \\ &= 1. \end{aligned}$$

下面我们证明, 对所有 y 成立

$$g(y) \geqslant f(y). \tag{4.18}$$

为此, 利用 (4.17) 有

$$g''(z) = 2a = 2\alpha f'(z) > f''(z).$$

再由 $g(z) = f(z)$ 和 $g'(z) = f'(z)$ 推知 (4.18) 在 z 的某个邻域成立. 于是 (4.18) 成立的充要条件是对所有 $y \leqslant l$,

$$h(y) = \frac{g(y)}{f(y)} - 1 \geqslant 0.$$

276

但 $h(y)$ 是三次式, 至多允许有三个实数根. 其中两个落在切点 z 上, 而第三个是 l. 由于 (4.18) 在 z 的邻域成立, 则对于 $y \leqslant l$ 必成立 $g(y) \geqslant f(y)$. 这样 (4.18) 对所有的 y 都成立. 有了这些预备知识, 我们转向证明 $f(x + S_n)$ 构成一个上鞅. 令 X 是随机变量, 均值为 m, 方差为 v^2, 且满足条件

$$m \leqslant -\alpha v^2. \tag{4.19}$$

令 $\xi \leqslant l$ 为一任意点, 并应用前面分析于 $z = \xi + m < l$(注意, 由 (4.19), m 是负的). 利用 (4.18) 我们有

$$\begin{aligned}E[f(\xi+X)] &= E[f(z+X-m)] \\ &\leqslant E[g(z+X-m)] \\ &= f(z)+\alpha f'(z)v^2 \\ &\leqslant f(z)-mf'(z) \qquad (\text{由于 } f'(z)>0 \text{及 } (4.19)) \\ &\leqslant f(z-m) \qquad (\text{由于当 } y<l \text{ 时有 } f''(y)\geqslant 0) \\ &= f(\xi).\end{aligned}$$

这样, 若 X 是满足 (4.19) 的随机变量, 对任意 $\xi \leqslant l$ 必成立 $f(\xi) \geqslant E[f(\xi+X)]$(当 $\xi > l$ 时是平凡的, 因为处处有 $f(z) \leqslant 1$). 由此结果, 并注意到 (4.13), 易推得

$$\begin{aligned}E[f(x+S_{n+1})|X_1,\cdots,X_n] &= E[f(x+S_n+X_{n+1})|X_1,\cdots,X_n] \\ &\leqslant f(x+S_n),\end{aligned}$$

这就验证了上鞅不等式. 因此 (4.14) 得证.

下面是 (4.14) 的等价形式, 由它可得许多重要的不等式.

设 $X_1', X_2', \cdots$ 是具有有限条件矩

$$M_k = E[X_k'|X_1',\cdots,X_{k-1}']$$

和

$$V_k = E[(X_k'-M_k)^2|X_1',\cdots,X_{k-1}']$$

的联合分布随机变量序列. 则

$$\begin{aligned}\Pr\{X_1'+\cdots+X_n'-(M_1+\cdots+M_n) \geqslant a(V_1+\cdots+V_n)+b \quad \text{对某个} n \geqslant 1\} \\ \leqslant \frac{1}{1+ab}, \qquad a,b \geqslant 0.\end{aligned}$$

277

此结果等价于

$$\Pr\{X_1+\cdots+X_n \geqslant b, \text{ 对某个} n \geqslant 1\} \leqslant \frac{1}{1+ab},$$

其中

$$X_k = X_n' - M_k - aV_k,$$

X_k 条件均值为 $-aV_k$, 而条件方差仍为 V_k, 这样新的不等式立刻由原不等式推得.

6.5　鞅收敛定理

在十分一般条件下, 当 n 增加时, 鞅 X_n 收敛于极限随机变量 X. 这些结果的精确论述构成了概率论中一些影响深远且强有力的定理. 我们下面集中精力于**鞅的基本收敛定理**. (第 1 章已回顾了随机变量序列几种收敛的概念.)

定理 5.1 (a) 设 $\{X_n\}$ 是下鞅, 满足

$$\sup_{n\geqslant 0} E[\,|X_n|\,] < \infty. \tag{5.1}$$

则存在一个随机变量 X_∞ 使 $\{X_n\}$ 以概率 1 收敛于 X_∞, 即

$$\Pr\left\{\lim_{n\to\infty} X_n = X_\infty\right\} = 1. \tag{5.2}$$

(b) 如果 $\{X_n\}$ 是鞅并且是一致可积的 (见后面附注 5.3), 则除成立 (5.2) 之外, $\{X_n\}$ 平均收敛, 即

$$\lim_{n\to\infty} E[\,|X_n - X_\infty|\,] = 0, \tag{5.3}$$

且对所有 n.

$$E[X_\infty] = E[X_n],$$

附注 5.1 对于一个下鞅 $\{X_n\}$, 如果 $E[\,|X_0|\,] < \infty$ 则条件

$$\sup_{n\geqslant 1} E[\,|X_n|\,] < \infty$$

278

等价于条件

$$\sup_{n\geqslant 1} E[X_n^+] < \infty,$$

其中 $X^+ = \max\{X, 0\}$. 这个等价性可由初等不等式 $X_n^+ \leqslant |X_n|$ 和关系式 $|X_n| = 2X_n^+ - X_n$ 得到, 只要注意到

$$E[\,|X_n|\,] = 2E[X_n^+] - E[X_n] \leqslant 2E[X_n^+] - E[X_0].$$

这个定理特别告诉我们, **每个非正下鞅以概率 1 收敛, 非负上鞅以及有一致上界或一致下界的鞅也同样如此**.

附注 5.2 以概率 1 收敛 (5.2), 并不能推得平均收敛 (5.3), 反过来也不行. 然而这两种收敛确实都蕴涵着依概率收敛

$$\lim_{n\to\infty} \Pr\{\,|X_n - X_\infty| > \varepsilon\,\} = 0, \qquad \text{对任意 } \varepsilon > 0.$$

事实上, 切比雪夫不等式 (参考 1.1 节)

$$\Pr\{\,|X_n - X_\infty| \geqslant \varepsilon\,\} \leqslant \frac{E[\,|X_n - X_\infty|\,]}{\varepsilon}, \quad \varepsilon > 0$$

表明平均收敛蕴涵着依概率收敛.

附注 5.3 由 6.3 节, 按定义, 如果

$$\lim_{c\to\infty}\sup_{n\geqslant 0} E[\,|X_n|\; I\{\,|X_n|>c\,\}]=0, \tag{5.4}$$

则 $\{X_n\}$ 一致可积. (记号 $I\{\,|X_n|>c\,\}$ 表示事件 $\{|X_n|>c\}$ 的示性函数). 特别地,

$$I\{\,|X_n|>c\,\}=\begin{cases}1, & \text{如果}|X_n|>c,\\ 0, & \text{如果}|X_n|\leqslant c.\end{cases}$$

今后, 记号 "I" 一般指事件的对应示性函数, 我们也把 $I\{\,|X_n|>c\,\}$ 写为 (见 6.2 节)$I_{\{\,|X_n|>c\,\}}$.

关系式 (5.4) 可由下面两条件之一推得:

(i) $|X_n|\leqslant W$, 对所有 n, (5.5)

其中 W 是满足 $E[W]<\infty$ 的随机变量;

(ii) $E[\,|X_n|^{1+\rho}\,]\leqslant K<\infty$, 对所有$n$, (5.6)

279 其中 K 和 ρ 是常数, $\rho>0$. (请读者证之).

我们将不证明一般形式的定理 5.1, 因为它要求许多测度论的知识. 我们推荐读者参看 Doob 的书 [1953, 第 7 章] 或 Neven 的书 [1965]. 上一节的上穿不等式是关键的工具. 然而, 我们将证明一个收敛性定理, 它与一般形式的证明相平行但所涉及的概念却比较简单. 特别地, 收敛定理 5.2 涉及满足较强条件的鞅, 即当 $\rho=1$ 时的条件 (5.6). 在这个比较强的条件下, 我们可得到更强的结论: 除了 X_n 以概率 1 收敛于 X_∞ 之外, 还以平方平均收敛于它. 下面讨论的最大值不等式是解决此问题的基本工具, 同时它还有许多其他用途.

最大值不等式

令 $\xi_1,\xi_2,\xi_3,\cdots$ 是独立同分布随机变量, 且满足矩条件 $E[\xi_i]=0$ 和 $E[\xi_i^2]=\sigma^2<\infty$. 定义

$$S_0=0 \text{ 且 } S_n=\xi_1+\cdots+\xi_n,\quad n\geqslant 1.$$

注意到 S_n 的方差是 $n\sigma^2$, 切比雪夫不等式给出

$$\varepsilon^2\Pr\{\,|S_n|>\varepsilon\,\}\leqslant n\sigma^2,\quad \varepsilon>0.$$

另一个更精致的不等式是科尔莫戈罗夫不等式

$$\varepsilon^2\Pr\Big\{\max_{0\leqslant k\leqslant n}|S_k|>\varepsilon\Big\}\leqslant n\sigma^2. \tag{5.7}$$

上述这些不等式容易推广到下鞅的情况, 这就是简单但却实用的**下鞅最大值不等式**(进一步加强的此不等式见问题 5).

引理 5.1 设 $\{X_n\}$ 是下鞅, 且对所有 n 有 $X_n \geqslant 0$. 则对于任意正数 λ, 有

$$\lambda \Pr\left\{\max_{0\leqslant k\leqslant n} X_k > \lambda\right\} \leqslant E[X_n]. \tag{5.8}$$

证明 定义马尔可夫时间

$$T = \begin{cases} \min\{k \geqslant 0; X_k > \lambda\}, & \text{如果 } X_k > \lambda \text{ 对某些 } k = 0, \cdots, n, \\ n, & \text{如果 } X_k \leqslant \lambda \text{ 对于 } k = 0, 1, \cdots, n. \end{cases}$$

280

应用类似于引理 3.2 的下鞅可选停止定理, 得到

$$\begin{aligned} E[X_n] &\geqslant E[X_T] \\ &\geqslant E\left[X_T\, I\left\{\max_{0\leqslant k\leqslant n} X_k > \lambda\right\}\right] \quad (\text{因为 } X_i \text{ 全部非负}) \\ &\geqslant \lambda \Pr\left\{\max_{0\leqslant k\leqslant n} X_k > \lambda\right\} \quad (\text{由于在所示集合上 } X_T \geqslant \lambda), \end{aligned}$$

此即所希望的不等式. ■

由于 $\{S_n\}$ 是一个鞅 [6.1 节例 (a)], 并依照引理 2.1, $X_n = S_n^2$ 确定一个非负下鞅, 由此即得科尔莫戈罗夫不等式.

$$\begin{aligned} n\sigma^2 &= E[S_n^2] \\ &\geqslant \lambda \Pr\left\{\max_{0\leqslant k\leqslant n} S_k^2 > \lambda\right\} \\ &= \varepsilon^2 \Pr\left\{\max_{0\leqslant k\leqslant n} |S_k| > \varepsilon\right\}. \quad \text{对于}\varepsilon = \sqrt{\lambda}. \end{aligned}$$

推论 5.1 设 $\{X_n\}$ 是鞅, 则对每个正数 λ

$$\lambda \Pr\left\{\max_{0\leqslant k\leqslant n} |X_k| > \lambda\right\} \leqslant E[\,|X_n|\,].$$

证明 如果 $\{X_n\}$ 是鞅, 由引理 3.1 知 $\{|X_n|\}$ 是非负下鞅. 然后利用刚刚证明的最大值不等式. ■

引理 5.1 的证明方法也适合于证明**上鞅的最大值不等式**, 其命题可叙述如下 (见问题 12):

引理 5.2 如果 $\{X_n\}$ 是非负上鞅, 则

$$\lambda \Pr\left\{\max_{0\leqslant k\leqslant n} X_k > \lambda\right\} \leqslant E[X_0], \text{对于}\lambda > 0.$$

例 定义 $X_0 = 1$ 和 $X_n = \prod_{i=1}^{n} Y_i$ 对于 $n \geqslant 1$, 其中 $Y_1, Y_2, \cdots$ 是非负独立随机变量且具有相同的均值 $E[Y_i] = 1$. 那么 $\{X_n\}$ 是非负鞅, 并且由最大值不等式知

281

$$\Pr\left\{\max_{0 \leqslant k \leqslant n} X_k > \lambda\right\} \leqslant 1/\lambda, \qquad \text{对于}\lambda > 0. \tag{*}$$

这个界是无甚俾益的, 我们可以解释如下. 考虑一个赌徒在每次丢一个硬币时, 以他拥有财富的 q 部分下赌, $0 < q < 1$, 假设赌徒的财富开始时刻是 1 元, 掷 n 次之后, 他的财富 X_n 是 $\prod_{j=1}^{n}(1+\delta_j q)$, 其中 δ_j 是一列独立随机变量, 可能值为 ± 1, 每个值出现的概率为 $\frac{1}{2}$. 因此, 如果赌徒有足够的耐心, 他将看到自己的财富减少到 0. 因为 $\{X_n\}$ 是一个鞅, 它当然是一个下鞅, 并且是正的,

$$\sup_{n \geqslant 1} E[\,|X_n|\,] = E[\,|X_n|\,] = 1,$$

所以 (5.1) 被满足. 这样, 当 $n \to \infty$ 时以概率 1 趋向于一个有限极限, 该极限一定是零, 因为每个其他的状态都是瞬态. 在这个问题中不等式 (*) 显得太弱.

鞅的均方收敛定理

设 $\{X_n\}$ 是鞅并令 A 是使得序列 $\{X_n\}$ 收敛的随机事件. 下面我们对集合 A 进行比较严格的描述. 注意到某个特定实现 $X_0, X_1, \cdots$ 收敛当且仅当满足柯西准则

$$\lim_{m,n \to \infty} |X_m - X_n| = 0,$$

于是 A 可以表达为

$$A = \left\{\lim_{m,n \to \infty} |X_m - X_n| = 0\right\}. \tag{5.9}$$

也就是说, A 是过程实现 $X_0, X_1, \cdots$ 满足柯西收敛准则的事件, 当 A 出现时, 令 X_∞ 表示这个极限. 我们要证明 $\Pr\{A\} = 1$. 则 X_∞ 可被定义, 虽然不是对所有实现 $X_0, X_1, \cdots$ 都有定义, 但至少对于具有概率为 1 的这种实现的集合有定义, 并且

$$\Pr\left\{\lim_{n \to \infty} X_n = X_\infty\right\} = 1.$$

在假设 $\{X_n\}$ 的二阶矩一致有界的情况下, 我们将证明它确实以概率 1 收敛并且均方收敛.

定理 5.2 设 $\{X_n\}$ 是关于 $\{Y_n\}$ 的鞅, 对于某个常数 K, 满足

282

$$E[X_n^2] \leqslant K < \infty, \qquad \text{对所有}n. \tag{5.10}$$

则当 $n \to \infty$ 时 $\{X_n\}$ 以概率 1 均方收敛于一个极限随机变量 X_∞, 即

$$\Pr\left\{\lim_{n\to\infty} X_n = X_\infty\right\} = 1 \tag{5.11}$$

和

$$\lim_{n\to\infty} E[\,|X_n - X_\infty|^2\,] = 0 \tag{5.12}$$

成立. 最后, 对于所有的 n

$$E[X_0] = E[X_n] = E[X_\infty]. \tag{5.13}$$

证明 暂时固定 N, 对于 $k \geqslant 0$ 置

$$\widetilde{X}_k = X_{N+k} - X_N.$$

由全概率定律得

$$\begin{aligned} E[X_{N+k}X_N] &= E\{E[X_{N+k}X_N|Y_0,\cdots,Y_N]\} \\ &= E\{X_N E[X_{N+k}|Y_0,\cdots,Y_N]\} \\ &= E[X_N^2]. \end{aligned}$$

所以

$$\begin{aligned} 0 &< E[\widetilde{X}_k^2] = E[(X_{N+k} - X_N)^2] \\ &= E[X_{N+k}^2 - 2X_{N+k}X_N + X_N^2] \\ &= E[X_{N+k}^2] - E[X_N^2]. \end{aligned} \tag{5.14}$$

引理 2.1 告诉我们, 当 $\{X_n\}$ 是鞅时 $\{X_n^2\}$ 是下鞅, 而 (5.14) 指出 $E[X_n^2]$ 是单调不减序列且上界为 K, 因此 $E[X_n^2]$ 收敛. 使用柯西准则得到

$$\begin{aligned} 0 &= \lim_{N,k\to\infty} \{E[X_{N+k}^2] - E[X_N^2]\} \\ &= \lim_{N,k\to\infty} E[\widetilde{X}_k^2]. \end{aligned} \tag{5.15}$$

下面我们马上就要用到这个结论.

设 A 是 $\{X_n\}$ 收敛的事件. 显然

$A =$ 满足 $\lim\limits_{m,n\to\infty} |X_m - X_n| = 0$ 的所有实现的集合

$= \{X_n\}$ 如下实现的集合：对任给$\varepsilon > 0$,

存在 $N > 0$,满足 $|X_{N+m} - X_{N+n}| \leqslant \varepsilon$,对所有$m, n \geqslant 1$.

由三角不等式

$$|X_{N+m}-X_{N+n}| \leqslant |X_{N+m}-X_N| + |X_{N+n}-X_N|,$$

我们可等价地把 A 描述为

$$A=\{\text{ 对任意}\varepsilon>0,\text{ 存在}N\geqslant 0,\text{使得对所有}k\geqslant 0\text{ 成立 }|X_{N+k}-X_N|\leqslant\varepsilon\ \}.$$

设 B 表示 A 的余事件, 即 $\{X_n\}$ 不收敛的集合, 则

$$\begin{aligned}B&=\{\text{对某个}\varepsilon>0,\text{对每个}N\geqslant 0,\text{存在}k\geqslant 0,k\text{取决于}\varepsilon\text{和}N,\text{使得}|X_{N+k}-X_N|>\varepsilon\}\\&=\bigcup_{\varepsilon>0}\{\text{对每个}N\geqslant 0,\text{存在 }k\geqslant 0,\text{使得 }|X_{N+k}-X_N|>\varepsilon\}\\&=\bigcup_{\varepsilon>0}\bigcap_{N=0}^{\infty}B_N(\varepsilon),\end{aligned}$$

其中

$$B_N(\varepsilon)=\text{由条件}\{\ |X_{N+k}-X_N|>\varepsilon\text{ 对于某个 }k=0,1,\cdots\}\text{描述的事件}.$$

我们希望证明等式 $\Pr\{B\}=0$. 为此只要证明

$$\lim_{N\to\infty}\Pr\{B_N(\varepsilon)\}=0,\quad\text{对每个}\varepsilon>0.\tag{5.16}$$

因为在这种情况下,

$$\begin{aligned}\Pr\{B\}&=\Pr\Big\{\bigcup_{\varepsilon>0}\bigcap_{N=0}^{\infty}B_N(\varepsilon)\Big\}\\&=\lim_{\varepsilon\downarrow 0}\Pr\Big\{\bigcap_{N=0}^{\infty}B_N(\varepsilon)\Big\}\\&=\lim_{\varepsilon\downarrow 0}\lim_{N\to\infty}\Pr\{B_N(\varepsilon)\}=0.\end{aligned}$$

为了证明 (5.16), 我们使用最大值不等式. 固定 N 并置

$$\widetilde{X}_k=X_{N+k}-X_N,\quad k=0,1,\cdots,$$

$$\widetilde{Y}_0=(Y_0,\cdots,Y_N),$$

且

284

$$\widetilde{Y}_k=Y_{N+k},\quad k=1,2,\cdots.$$

由 Jensen 不等式, 或施瓦兹不等式, 并利用 (5.14), 我们得到

$$E[\,|\widetilde{X}_k|\,] \leqslant (E[\widetilde{X}_k^2])^{1/2} \leqslant \sqrt{K} < \infty$$

和

$$\begin{aligned} E[\widetilde{X}_{k+1}|\widetilde{Y}_0,\cdots,\widetilde{Y}_k] &= E[X_{N+k+1}-X_N|Y_0,\cdots,Y_{N+k}] \\ &= X_{N+k}-E[X_N|Y_0,\cdots,Y_{N+k}] \\ &= X_{N+k}-X_N=\widetilde{X}_k, \end{aligned}$$

所以 $\{\widetilde{X}_k\}$ 是关于 $\{\widetilde{Y}_k\}$ 的鞅. 此外, 引理 2.1 告诉我们 $\{\widetilde{X}_k^2\}$ 是下鞅, 并且由最大值不等式得

$$\varepsilon^2 \Pr\Big\{\max_{0\leqslant k\leqslant n}|\widetilde{X}_k^2|>\varepsilon^2\Big\} \leqslant E[\widetilde{X}_n^2],$$

或

$$\varepsilon^2 \Pr\Big\{\max_{0\leqslant k\leqslant n}|X_{N+k}-X_N|>\varepsilon\Big\} \leqslant E[\widetilde{X}_n^2].$$

但在 (5.15) 我们证明了, 当 $N,n\to\infty$ 时其右边部分趋于零. 由此推得

$$\begin{aligned} 0 &= \lim_{N\to\infty}\lim_{n\to\infty}\Pr\Big\{\max_{0\leqslant k\leqslant n}|X_{N+k}-X_N|>\varepsilon\Big\} \\ &= \lim_{N\to\infty}\Pr\Big\{\sup_{0\leqslant k<\infty}|X_{N+k}-X_N|>\varepsilon\Big\} \\ &= \lim_{N\to\infty}\Pr\{B_N(\varepsilon)\}. \end{aligned}$$

这就证明了 $\{X_n\}$ 以概率 1 收敛.

用 X_∞ 表示极限随机变量. 尚需证明 $\{X_n\}$ 均方收敛于 X_∞. 我们仅需证明下面第二个不等号即可:

$$\begin{aligned} 0 &\leqslant \lim_{n\to\infty}E[\,|X_n-X_\infty|^2\,] \\ &= \lim_{n\to\infty}E\Big[\lim_{m\to\infty}|X_n-X_m|^2\Big] \\ &\leqslant \lim_{n\to\infty}\lim_{m\to\infty}E[\,|X_n-X_m|^2\,]=0. \end{aligned}$$

最后式子的极限为 0 是由于 (5.15). 固定 n, 令 $Z_m=|X_n-X_m|^2$, 所以我们需要证明

$$\lim_{m\to\infty}E[Z_m] \geqslant E\Big[\lim_{m\to\infty}Z_m\Big]. \tag{5.17}$$

285

首先, 我们有

$$\begin{aligned}E[Z_m] &= \int_0^{\infty} \Pr\{Z_m \geqslant t\}\mathrm{d}t \\ &= \sum_{k=1}^{\infty} \int_{(k-1)\varepsilon}^{k\varepsilon} \Pr\{Z_m \geqslant t\}\mathrm{d}t \\ &\geqslant \sum_{k=1}^{\infty} \int_{(k-1)\varepsilon}^{k\varepsilon} \Pr\{Z_m \geqslant k\varepsilon\}\mathrm{d}t \\ &\geqslant \sum_{k=1}^{N} \varepsilon \Pr\{Z_m \geqslant k\varepsilon\},\end{aligned} \tag{5.18}$$

其中 $\varepsilon > 0, N > 0$ 是任意的.

由附注 5.2 知, 以概率 1 收敛可推得依概率收敛, 由此, 对于 $\delta > 0$,

$$\lim_{m\to\infty} \Pr\{Z_m \geqslant k\varepsilon\} \geqslant \lim_{m\to\infty} [\Pr\{Z \geqslant k\varepsilon + \delta\} - \Pr\{\ |Z_m - Z| > \delta\ \}] = \Pr\{Z \geqslant k\varepsilon + \delta\},$$

其中 $Z = \lim_{m\to\infty} Z_m$.

由于 $\delta > 0$ 是任意的,

$$\lim_{m\to\infty} \Pr\{Z_m \geqslant k\varepsilon\} \geqslant \Pr\{Z > k\varepsilon\}.$$

现在回到 (5.18), 我们有

$$\begin{aligned}\lim_{m\to\infty} E[Z_m] &\geqslant \lim_{m\to\infty} \sum_{k=1}^{N} \varepsilon \Pr\{Z_m \geqslant k\varepsilon\} \\ &\geqslant \sum_{k=1}^{N} \varepsilon \Pr\{Z > k\varepsilon\} \\ &\geqslant \sum_{k=0}^{N} \varepsilon \Pr\{Z > k\varepsilon\} - \varepsilon \\ &\geqslant \sum_{k=0}^{N} \int_{k\varepsilon}^{(k+1)\varepsilon} \Pr\{Z > t\}\mathrm{d}t - \varepsilon \\ &\geqslant \int_0^{(N+1)\varepsilon} \Pr\{Z > t\}\mathrm{d}t - \varepsilon\end{aligned}$$

保持 $\varepsilon > 0$ 不变, 令 $N \to \infty$, 导出

286

$$\lim_{m\to\infty} E[Z_m] \geqslant \int_0^{\infty} \Pr\{Z > t\}\mathrm{d}t - \varepsilon = E[Z] - \varepsilon.$$

由于 $\varepsilon > 0$ 是任意的, (5.17) 得证, 因此 $\{X_n\}$ 均方收敛于 X_∞. 依施瓦兹不等式, 这蕴涵着平均收敛, 即

$$0 \leqslant E[\,|X_n - X_\infty|\,] \leqslant \{E[\,|X_n - X_\infty|^2\,]\}^{1/2} \to 0, \text{当} n \to \infty.$$

另一方面, 我们有

$$\begin{aligned} 0 &\leqslant |E[X_n] - E[X_\infty]| \\ &= |E[X_n - X_\infty]| \leqslant E[\,|X_n - X_\infty|\,] \to 0, \text{当} n \to \infty, \end{aligned}$$

所以, $E[X_n] \to E[X_\infty]$. 但 $E[X_n] = E[X_0]$ 对于所有的 n 成立, 故

$$E[X_0] = E[X_n] = E[X_\infty],$$

这样就证明了鞅的均方收敛定理.

6.6 鞅收敛定理的应用和扩展

下面是鞅收敛定理的若干应用实例.

(a) $y = Py$ 的有界解, 其中 $P = \|P_{ij}\|$ 是不可经常返马尔可夫链 $\{Y_n\}$ 的转移矩阵. 鞅的收敛定理可用于证明.

$$y(i) = \sum_{j=0}^{\infty} P_{ij} y(j)$$

的有界解 $y = \{y(i)\}$ 对于所有的 i 是一个常数, 即对于所有的 $i, j, y(i) = y(j)$ 成立. (与第 3 章定理 4.1 比较). 因为 $X_n = y(Y_n)$ 是有界鞅 [参考 6.1 节例 (d)], 所以以概率 1 存在 $\lim\limits_{n\to\infty} X_n = \lim\limits_{n\to\infty} y(Y_n)$. 由于链是常返的, 所以所有状态都将无限多次达到 (见第 2 章), 因此

$$\{X_n = y(i)\} \quad \text{和} \quad \{X_n = y(j)\}$$

一定都出现无限多次. 但 $\lim\limits_{n\to\infty} X_n$ 存在, 于是我们必有 $y(i) = y(j)$.

(b) $f(y) = \int f(y+z)p(z)\mathrm{d}z$ 的解. 设 $p(z)$ 是连续的概率密度函数. 每个常数函数 $f(y) \equiv a$ 都是积分方程

$$f(y) = \int f(y+z)p(z)\mathrm{d}z \qquad \text{对于所有的} y \tag{6.1}$$

287

的一个解. 鞅收敛定理能够证明 (6.1) 的唯一**有界连续解**一定是常数函数. 为此假设 $f(y)$ 是有界连续的, 并且是 (6.1) 的解. 那么对每个 $x, \{f(x + S_n)\}$ 构成一个鞅序

列, 其中 $\{S_n\}$ 表示由独立同分布随机变量 $X_1, X_2, \cdots$ 产生的部分和序列, X_i 的公共概率密度函数是 p. 由于 $f(y)$ 是有界的, 满足均方收敛定理的条件, 我们断定对每个 x 存在随机变量 U_x 满足.

$$\Pr\{\lim_{n\to\infty} f(x+S_n) = U_x\} = 1$$

且

$$\lim_{n\to\infty} E[\,|f(x+S_n) - U_x|^2\,] = 0.$$

我们首先证明 U_x 不是随机的, 而是一个常数 $U_x = u(x)$, 然后我们将证明 $u(x)$ 实际上与 x 无关.

显然, $f(x+S_m-S_n)$ 与 $f(x+S_{m-n})$ 具有相同的分布, 其中 $m \geqslant n$. 比较隐蔽的性质是二维随机向量

$$\{f(x+S_m), \quad f(x+S_m-S_n)\}$$

与二维随机向量

$$\{f(x+S_m), f(x+S_{m-n})\}, m \geqslant n.$$

具有相同的联合分布 (为什么?). 利用这些事实, 我们有

$$\begin{aligned} &\lim_{n\to\infty}\lim_{m\to\infty} E[\,|f(x+S_m) - f(x+S_m-S_n)|^2\,] \\ =\ &\lim_{n\to\infty}\lim_{m\to\infty} E[\,|f(x+S_m) - f(x+S_{m-n})|^2\,] = 0. \end{aligned} \tag{6.2}$$

这是极为重要的一步, 下面我们立刻会看到这一点.

根据施瓦兹不等式, 我们有

$$\begin{aligned} \{E[U_x^2 - f(x+S_n)U_x]\}^2 &= \{E[U_x\{U_x - f(x+S_n)\}]\}^2 \\ &\leqslant E[U_x^2]\cdot E[\,|U_x - f(x+S_n)|^2\,]. \end{aligned}$$

但是由当 n 增加时, 右端趋于 0, 可知

$$E[U_x^2] = \lim_{n\to\infty} E[f(x+S_n)U_x].$$

类似地, 我们可导出

288

$$E[f(x+S_n)U_x] = \lim_{m\to\infty} E[f(x+S_n)f(x+S_m)].$$

综合这些关系式得到

$$\begin{aligned} E[U_x^2] &= \lim_{n\to\infty}\lim_{m\to\infty} E[f(x+S_n)f(x+S_m)] \\ &= \lim_{n\to\infty}\lim_{m\to\infty} E[f(x+S_n)f(x+S_m-S_n)] \\ &\quad + \lim_{n\to\infty}\lim_{m\to\infty} E[f(x+S_n)\{f(x+S_m) - f(x+S_m-S_n)\}]. \end{aligned}$$

对于右边第一项由于 S_n 和 $S_m - S_n$ 是独立的, 并且对于所有的 n 有 $E[f(x+S_n)] = E[U_x]$, 由鞅收敛定理, 我们有

$$\begin{aligned} &E[f(x+S_n)f(x+S_m-S_n)] \\ &= E[f(x+S_n)]E[f(x+S_m-S_n)] \\ &= \{E[U_x]\}^2. \end{aligned}$$

考察右端第二项, 利用施瓦兹不等式, 我们得到

$$\begin{aligned} &\{E[f(x+S_n)\{f(x+S_m)-f(x+S_m-S_n)\}]\}^2 \\ &\leqslant E[\{f(x+S_n)\}^2]\cdot E[\ |f(x+S_m)-f(x+S_m-S_n)|^2\ ], \end{aligned}$$

由于 $f(y)$ 是有界的, 当 $m\to\infty$ 时最后的因子趋于 0, 这里使用了重要关系式 (6.2).

综合上述关系, 导出等式

$$E[U_x^2] = \{E[U_x]\}^2,$$

即 U_x 的方差是 0, 这意味着 U_x 是一个非随机常数, 记为 $u(x)$.

鞅具有常数均值, 利用鞅的均方收敛定理可知, 此均值也是极限随机变量的均值, 因此

$$f(x) = E[f(x+S_0)] = E[U_x] = u(x).$$

至此我们已证明了以概率 1,

$$\lim_{n\to\infty} f(x+S_n) = u(x) = f(x).$$

但我们有

$$\lim_{n\to\infty} f(x+X_1+S_n-X_1) = f(x+X_1),$$

所以

$$\Pr\{f(x) = f(x+X_1)\} = 1,$$

并由归纳法得

$$\Pr\{f(x) = f(x+S_n)\} = 1.$$

289

利用这个事实以及假设 X_1 是具有概率密度函数 $p(x)$ 的连续型随机变量, 我们可以证明对于任意 x, y 有 $f(x) = f(y)$. 例如当 $p(z) > 0$ 对于所有的 z 成立时, 立即可得到此结论, 因为

$$0 = E[|f(x+X_1)-f(x)|] = \int |f(x+z)-f(x)|p(z)\mathrm{d}z.$$

因此对于所有的 z, $|f(x+z)-f(x)|=0$, 这里利用了假设 $f(x)$ 是连续的.

由此推得对于所有的 x, $f(x)=a$, 即 (6.1) 的每一个连续解是常数.

如果假设 $f(x)$ 在点 x_0 达到其最大值, 证明可简单许多. 在这个附加假设之下, 由 (6.1) 得到

$$0=\int\{f(x_0)-f(x_0+y)\}p(y)\mathrm{d}y=\int|f(x_0)-f(x_0+y)|p(y)\mathrm{d}y,$$

当假设对于所有的 y 有 $p(y)>0$ 时, 容易推得对于所有的 y 成立 $f(x_0)=f(x_0+y)$. 然而, 这样简单的证明对一般情况是行不通的.

(c) 罐模型. 考虑 6.1 节例 (i) 提出的罐模型, 初时罐中装有 n 个红球和 m 个绿球, 如前述, 用 X_k 表示第 $k(k=0,1,2,\cdots)$ 步罐中红球的比例数, 则 $\{X_k\}$ 是一个有界鞅, 所以 $X_\infty=\lim\limits_{k\to\infty}X_k$ 以概率 1 存在.

极限随机变量 X_∞ 具有 Beta 分布. 我们简单推导如下. 设 $Y_k=(n+m+k)X_k$ 是第 k 步罐中红球的数目. 关于 k 使用归纳法推得公式

$$\Pr\{Y_k=i\}=\frac{\begin{pmatrix}i-1\\n-1\end{pmatrix}\begin{pmatrix}N-i-1\\m-1\end{pmatrix}}{\begin{pmatrix}N-1\\n+m-1\end{pmatrix}},\quad n\leqslant i\leqslant n+k,\ N=n+m+k.$$

利用斯特林近似公式 $M!\sim\mathrm{e}^{-M}M^M(2\pi M)^{1/2}$, 对于比较大的 M 和固定的 j 可得 $\begin{pmatrix}M\\j\end{pmatrix}\sim(M-j)^j/j!$. 则

$$\begin{aligned}\Pr\{X_k\leqslant x\}&=\Pr\{Y_k\leqslant Nx\}\\&=\sum_{i=0}^{[Nx]}\begin{pmatrix}i-1\\n-1\end{pmatrix}\begin{pmatrix}N-i-1\\m-1\end{pmatrix}\Bigg/\begin{pmatrix}N-1\\n+m-1\end{pmatrix}\\&\approx\frac{(n+m-1)!}{(n-1)!(m-1)!}\sum_{i=0}^{[Nx]}\left(\frac{i-n}{k}\right)^{n-1}\left(\frac{k-i+n}{k}\right)^{m-1}\frac{1}{k}\\&=\frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)}\sum_{i/k=0}^{[Nx]/k}\left(\frac{i}{k}-\frac{n}{k}\right)^{n-1}\left(1-\frac{i}{k}+\frac{n}{k}\right)^{m-1}\Delta\left(\frac{i}{k}\right),\end{aligned}$$

290

其中 $\Delta(i/k)=(i+1)/k-i/k=\dfrac{1}{k}$. 当 N 和 k 较大时, n/k 可以忽略, 求和可以用

积分号来代替, 其中, $z=i/k, \mathrm{d}z \approx \Delta(i/k)$, 并且 $[Nx]/k \sim x$. 这样, 当 $k \to \infty$ 时,

$$\begin{aligned}\Pr\{X_k \leqslant x\} &\to \frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)} \int_0^x z^{n-1}(1-z)^{m-1}\mathrm{d}z \\ &= \Pr\{X_\infty \leqslant x\}, \qquad \text{对于 } 0 \leqslant x \leqslant 1.\end{aligned}$$

这就是所要证明的 Beta 分布.

(d) 分支过程. 6.1 节例 (f) 指出 $X_n = m^{-n}Y_n$ 是鞅, 其中 $m < \infty$ 是分支过程 $\{Y_n\}$ 后代分布的均值. 由附注 5.1 所说明的基本鞅收敛定理告诉我们, $X_\infty = \lim_{n\to\infty} X_n$ 以概率 1 存在. 粗略地说, Y_n 的性质渐近于 $X_\infty m^n$, 值得注意的是整个渐近性质可由单一的随机变量 X_∞ 获得.

在第 8 章我们将证明, 只要 $m \leqslant 1, Y_n \to 0$ 以概率 1 成立. 因为 Y_n 的可能值是 $\{0,1,\cdots\}$, 这意味着对于过程几乎所有的实现, 当 n 足够大时 $Y_n=0$, 因此

$$X_\infty = \lim_{n\to\infty} m^{-n}Y_n = 0.$$

这样, 由于

$$1 = E[m^{-n}Y_n] \neq E[X_\infty] = 0,$$

所以当 $m \leqslant 1$ 时序列 $X_n = m^{-n}Y_n$ 不可能一致可积.

当 $m > 1$ 时, 后代分布具有有限方差 σ^2, 由 8.2 节, 我们得到方差

$$\mathrm{Var}[Y_n] = \frac{\sigma^2}{m}\left(\frac{m^{2n}-m^n}{m-1}\right),$$

利用这个式子得到

$$\begin{aligned}E[X_n^2] &= \{\mathrm{Var}[Y_n] + (E[Y_n])^2\}/m^{2n} \\ &= \frac{\sigma^2}{m}\left(\frac{1-m^{-n}}{m-1}\right) + 1 \\ &\leqslant \frac{\sigma^2}{m(m-1)} + 1, \qquad \text{对于所有的} n.\end{aligned}$$

由此可知, 满足鞅的均方收敛定理的条件. 这样, $1 = X_0 = E[X_\infty]$, 因此以正概率 X_∞ 是严格正的, 则 $Y_n \sim X_\infty m^n$ 表明 Y_n 的增长速度渐近于比率为 m 的指数增长速度.

291

(e) 分支过程中的分裂时间. 考虑由粒子构成的群体, 每个粒子具有独立随机寿命, 在生命结束时, 每个粒子就分裂成随机数目的新粒子, 新粒子各自独立地具有原来粒子相同的寿命特性. 特别地, 假设寿命时间是具有相同参数 a 的相互独立的指数分布随机变量.

设 $X(t)$ 是在时刻 t 群体粒子的数目, τ_n 是群体第 n 次分裂的时刻 $(\tau_0 = 0)$. 随机变量 $\xi_n = X(\tau_n + 0) - X(\tau_n - 0)$ 表示第 n 次分裂所增加的后代的数目. 设

$$X(0) = 1,$$

$$S_i = X(\tau_i) = X(\tau_i + 0) = \xi_1 + \xi_2 + \cdots + \xi_i + 1,$$

并定义 $T_i = \tau_i - \tau_{i-1}$ 是第 $(i-1)$ 次分裂和第 i 次分裂之间的时间, 我们断言随机变量序列

$$Y_n = \sum_{i=1}^{n} \left(T_i - \frac{1}{aS_{i-1}} \right), n = 1, 2, \cdots,$$

是关于 $\{(\tau_n, \xi_n)\}$ 的鞅. 我们只需要验证 [参考 6.1 节例 (b)]

$$E\left[T_n - \frac{1}{aS_{n-1}} \middle| \tau_0, \cdots, \tau_{n-1}, \xi_0, \cdots, \xi_{n-1}\right] = 0. \tag{6.3}$$

指数分布的无记忆性一些相关定义蕴涵着 T_n 是 S_{n-1} 个独立寿命的最小值, 每个独立寿命都服从参数为 a 的指数分布, 于是, T_n 是服从参数 aS_{n-1} 的指数分布 (参见第 4 章初等问题 1). 由此即可推得等式 (6.3).

进一步的计算也利用了和的鞅性质以及每个被加项的指数分布, 我们有 $Y_0 = 0$ 及

$$\begin{aligned}
E[Y_n^2] &= \sum_{k=1}^{n} E[(Y_k - Y_{k-1})^2] \qquad (\text{见问题 } 3) \\
&= \sum_{k=1}^{n} E\left[(T_k - \frac{1}{aS_{k-1}})^2\right] \\
&= \sum_{k=1}^{n} E\left\{E\left[\left(T_k - \frac{1}{aS_{k-1}}\right)^2 \middle| S_{k-1}\right]\right\} \\
&= \sum_{k=1}^{n} E\left[\left(\frac{1}{aS_{k-1}}\right)^2\right] \\
&\leqslant \frac{1}{a^2}\left\{1 + \sum_{k=1}^{\infty} \frac{1}{k^2} E\left[\left(\frac{k}{S_k}\right)^2\right]\right\}.
\end{aligned}$$

292

我们马上要证明 $E[(k/S_k)^2]$ 是一致有界的, 设其界为 C, 则

$$E[Y_n^2] \leqslant a^{-2}[1 + C\sum_{k=1}^{\infty} k^{-2}] < \infty, \quad n = 1, 2, \cdots,$$

因而验证了鞅均方收敛定理的条件.

为了得到界 C, 令 $V_k = \widetilde{\xi}_1 + \widetilde{\xi}_2 + \cdots + \widetilde{\xi}_k$, 其中

$$\widetilde{\xi}_i = \begin{cases} 0, & \text{如果 } \xi_i = 0, \\ 1, & \text{如果 } \xi_i \geqslant 1. \end{cases}$$

则 V_k 具有参数为 k 以及 $p=\Pr\{\xi_i>0\}>0$ 的二项分布, 并且我们知道

$$E[s^{V_k}]=[1-p+ps]^k, \qquad 0\leqslant s\leqslant 1,$$

及 $1+V_k\leqslant S_k$, 因此

$$\begin{aligned}E\Big[\frac{k}{S_k}\Big]&\leqslant kE\Big[\frac{1}{1+V_k}\Big]\\&=k\int_0^1 E[s^{V_k}]\mathrm{d}s\\&=k\int_0^1[1-p+ps]^k\mathrm{d}s\\&=[k/(k+1)][1-(1-p)^{k+1}]/p\\&\leqslant p^{-1}<\infty.\end{aligned}$$

同时注意到, 对于 $0<c\leqslant 1$ 有

$$E\Big[\frac{k}{c+\xi_1+\cdots+\xi_k}\Big]\leqslant\frac{1}{c}E\Big[\frac{k}{1+\xi_1+\cdots+\xi_k}\Big]\leqslant(cp)^{-1}<\infty.$$

293

其次, 利用初等不等式 $(a+b)^2\geqslant 4ab$ 和 $\xi_1,\xi_2,\cdots$ 的独立性, 得

$$\begin{aligned}E\left[\left(\frac{2k}{S_{2k}}\right)^2\right]&=E\left[\left\{\frac{2k}{\left(\frac{1}{2}+\xi_1+\cdots+\xi_k\right)+\left(\frac{1}{2}+\xi_{k+1}+\cdots+\xi_{2k}\right)}\right\}^2\right]\\&\leqslant E\left[\frac{k}{\frac{1}{2}+\xi_1+\cdots+\xi_k}\cdot\frac{k}{\frac{1}{2}+\xi_{k+1}+\cdots+\xi_{2k}}\right]\\&\leqslant\left\{E\left[\frac{k}{\frac{1}{2}+\xi_1+\cdots+\xi_k}\right]\right\}^2\leqslant\left(\frac{2}{p}\right)^2.\end{aligned}$$

即当 n 是偶数时, 它是 $E[(n/S_n)^2]$ 的界, 当 n 是奇数时,

$$\begin{aligned}E\Big[\Big(\frac{n+1}{S_{n+1}}\Big)^2\Big]&\leqslant\Big(\frac{n+1}{n}\Big)^2E\Big[\Big(\frac{n}{S_n}\Big)^2\Big]\\&\leqslant 4\Big(\frac{2}{p}\Big)^2.\end{aligned}$$

由是, 如果我们取 $C=16/p^2$, 则对于所有的 k 都有 $E[(k/S_k)^2]\leqslant C$.

我们应用鞅均方收敛定理得到 $Y_\infty=\lim\limits_{n\to\infty}Y_n$ 以概率 1 存在. 由于 $\sum\limits_{i=1}^n T_i=\tau_n$,

$$\tau_n-a^{-1}\sum_{i=1}^n S_{i-1}^{-1}\to Y_\infty, \quad 当\ n\to\infty. \tag{6.4}$$

我们分析级数 $\sum_{i=1}^{n} 1/(S_{i-1})$ 的性质之后可以得到进一步推论. 设 $\mu = E[\xi_i] > 0$. 由强大数定律我们知道

$$\frac{1}{n} S_n \to \mu, \qquad 当 n \to \infty,$$

以概率 1 成立, 我们写

$$\frac{1}{\ln n} \sum_{i=1}^{n} \frac{1}{S_{i-1}} = \frac{1}{\ln n} \sum_{i=1}^{N} \frac{1}{S_{i-1}} + \frac{1}{\ln n} \sum_{i=N}^{n-1} \frac{1}{i} \frac{i}{S_i}.$$

对任意 $\varepsilon > 0$, 选择 N 足够大使得对于所有的 $i \geqslant N, |i/S_i - \mu^{-1}| \leqslant \varepsilon$. 另一方面, 对任何固定的 N, 当 $n \to \infty$ 时上式右端的第一项趋于 0. 由这些估计导出结果:

294

$$\limsup_{n \to \infty} \frac{1}{\ln n} \left| \sum_{i=1}^{n} \frac{1}{S_{i-1}} - \sum_{i=N}^{n} \frac{1}{\mu i} \right| \leqslant \varepsilon,$$

因为 ε 是任意的, 我们得到以概率 1

$$\lim_{n \to \infty} \frac{1}{\ln n} \sum_{i=1}^{n} \frac{1}{S_{i-1}} = \frac{1}{\mu}. \tag{6.5}$$

结合极限 (6.4) 与 (6.5), 我们求得

$$\tau_n - \frac{\ln n}{a \mu} \to Y_\infty, \qquad 当 n \to \infty.$$

并且此极限也是在概率 1 的意义下取的. 另一方面, 值是注意的是

$$\tau_{2n} - \tau_n \cong (a\mu)^{-1} \ln 2$$

却渐近于一个常数

(f) Doob过程. 我们将说明 Doob 过程 [见 6.1 节例 (k)] 是一致可积的, 并且满足鞅收敛定理的所有条件. 令 $Z, Y_0, Y_1, \cdots$ 是一列随机变量且 $E[\,|Z|\,] < \infty$. 我们已经证明 (见 6.1 节例 (k))

$$X_n = E[Z | Y_0, \cdots, Y_n], \qquad n = 0, 1, 2, \cdots$$

确定了一个鞅, 并且满足

$$E[\,|X_n|\,] \leqslant E[\,|Z|\,], \qquad 对于所有的 n, \tag{6.6}$$

由鞅的最大值不等式得到

$$\Pr\left\{ \max_{0 \leqslant k \leqslant n} |X_k| > \lambda \right\} \leqslant \lambda^{-1} E[\,|X_n|\,] \leqslant \lambda^{-1} E[\,|Z|\,],$$

这样, 对于 $U = \sup_{k \geqslant 0} |X_k|$ 有

$$\Pr\{U > \lambda\} \leqslant \lambda^{-1} E[\,|Z|\,].$$

重要的是

$$\lim_{\lambda \to \infty} \Pr\{U > \lambda\} = 0, \tag{6.7}$$

推得 U 是有限值随机变量. 因此, 当 $N \to \infty$ 时有

$$\begin{aligned} E[\,|Z|\,] &\geqslant E[\,|Z| I\{0 \leqslant U < N\}] \qquad \text{(记号 } I\{\quad\}\text{见附注 5.3.)} \\ &= \sum_{k=0}^{N-1} E[\,|Z| \big| k \leqslant U < k+1] \Pr\{k \leqslant U < k+1\} \\ &\to \sum_{k=0}^{\infty} E[\,|Z| \big| k \leqslant U < k+1] \Pr\{k \leqslant U < k+1\} \\ &= E[\,|Z|\,], \end{aligned}$$

由此得

295

$$\lim_{N \to \infty} E[\,|Z|\; I\{U \geqslant N\}] = 0. \tag{6.8}$$

我们进一步验证附注 5.3 的一致可积性要求. 由 (6.8) 有

$$\begin{aligned} &|E[X_n I\{|X_n| > c\}]| \\ &\leqslant E[(I\{|X_n| > c\})(E[\,|Z| \big| Y_0, \cdots, Y_n])] \\ &\leqslant E[\,|Z| I\{\,|X_n| > c\}] \\ &\leqslant E[\,|Z| I\{U > c\}] \to 0, \qquad \text{当} c \to \infty. \end{aligned}$$

这样, $\{X_n\}$ 一致可积, 因此我们可以推得, 对某随机变量 X_∞,

$$\lim_{n \to \infty} X_n = X_\infty,$$

并且关系式

$$E[X_\infty] = E[X_0]$$

和

$$E[\,|X_n - X_\infty|\,] \to 0, \qquad \text{当 } n \to \infty$$

都成立. 关系式

$$X_\infty = E[Z | Y_0, Y_1, \cdots], \tag{6.9}$$

是正确的, 尽管这个条件期望的准确意义已超出本节的范围 (见 6.7 节).

下面将把上述结果应用于数学分析. 假设 W 是 $[0,1)$ 上均匀分布的随机变量, 定义 Y_n 如下:

$$Y_n = k2^{-n}, \qquad \text{对于 } k2^{-n} \leqslant W < (k+1)2^{-n}.$$

由 6.1 节例 (l) 知, $Y_0, \cdots, Y_n$ 确定了 W 的有尽二进位展开的前 n 位.

令 f 是定义在 $[0, 1]$ 上的任意函数满足

$$\int_0^1 |f(w)| \mathrm{d}w < \infty.$$

置 $Z = f(W)$ 并注意到

$$\begin{aligned} X_n &= E[Z|Y_0, \cdots, Y_n] \\ &= 2^n \int_{k/2^n}^{(k+1)/2^n} f(w)\mathrm{d}w, \qquad \text{如果 } k/2^n \leqslant W < (k+1)/2^n. \end{aligned}$$

296

由刚刚证明的关于 Doob 过程的结果知, $\lim\limits_{n\to\infty} X_n = X_\infty$ 存在, 并且由于 $Y_0, Y_1, Y_2, \cdots$ 给出 W 的完全二进位展开, 因此自然可以假设

$$\begin{aligned} X_\infty &= E[Z|Y_0, Y_1, \cdots] \\ &= E[f(W)|W] \\ &= f(W). \end{aligned}$$

虽然这已越出我们的范围, 但如果添上 "以概率 1" 的限制, 则它确实是正确的. 于是我们已证明了以概率 1,

$$f(W) = \lim_{n\to\infty} f_n(W),$$

此处

$$f_n(w) = 2^n \int_{k(w)/2^n}^{[k(w)+1]/2^n} f(z)\mathrm{d}z,$$

其中 $k(w)$ 由不等式

$$k(w)/2^n \leqslant w < [k(w)+1]/2^n.$$

唯一确定.

每个逼近函数 f_n 都是阶梯函数, 它在每个区间 $[k/2^n, (k+1)/2^n)$ 都是常数. 这样, 任意可积函数可以用一个阶梯函数序列 f_n 依下述意义来逼近; $f(w) = \lim\limits_{n\to\infty} f_n(w)$ 对于 w"几乎每个" 成立, 即对概率为 1 的集合中的每个 w 都成立.

最后, 由 $E[\,|f(W) - f_n(W)|\,] \to 0$ 可推得

$$\lim_{n\to\infty} \int_0^1 |f(z) - f_n(z)| \mathrm{d}z = 0.$$

6.7 关于 σ 域族的鞅

直到现在为止我们总认为条件期望是利用条件公布计算的期望. 这对于形如 $E[X|Y_0,\cdots,Y_n]$ 的条件期望大体上是可行的, 其中 $X,Y_0,\cdots,Y_n$ 具有联合密度函数或都是离散随机变量. 然而, 当推广到更加一般的情况, 例如 $E[X|Y_0,Y_1,\cdots]$ 或 $E[X|Y(u),0\leqslant u\leqslant t]$, 就难以处理了.

另一个比较现代的处理方法是定义和计算关于某个事件集合的条件期望, 这个集合被称为一个 σ 域, 而不再仅仅是关于有限个随机变量. 这样, 自然促使我们定义关于一个 σ 域族的鞅. 下面我们简要叙述一下这些在现代概率论著作中经常出现的概念.

297

概率公理化体系概述

本书大部分通过分布函数研究随机变量. 例如, 在一开始我们就认为一个随机过程 $\{X(t);\ t\in T\}$ 是确定的, 一旦所有的有限维分布

$$F(x_1,\cdots,x_n,t_1,\cdots,t_n)=\Pr\{X(t_1)\leqslant x_1,\cdots,X(t_n)\leqslant x_n\}$$

都确定. 现在我们需要更多的细节和结构.

回忆概率理论中基本元素是:

(1) **样本空间**, 即集合 Ω, 其元素 ω 对应一个实验的可能结果;

(2) **事件族**, 即 Ω 的某些子集的集合 $\mathcal{F}$. 如果实验结果 ω 是 A 的元素, 我们便说事件 A 出现;

(3) **概率测度**, 即定义在 $\mathcal{F}$ 上的函数 P, 它满足

(a) $0=P[\varnothing]\leqslant P[A]\leqslant P[\Omega]=1,\qquad A\in\mathcal{F},\qquad(\varnothing=\text{空集}),$

(b) $P[A_1\bigcup A_2]=P[A_1]+P[A_2]-P[A_1\bigcap A_2],\qquad A_i\in\mathcal{F},\qquad i=1,2,\quad(7.1)$

(c) $P\left[\bigcup_{n=1}^{\infty}A_n\right]=\sum_{n=1}^{\infty}P[A_n], A_n\in\mathcal{F}$互不相交,即 $A_i\bigcap A_j=\varnothing,\quad i\neq j.$

此三元组 $(\Omega,\mathcal{F},P)$ 称为**概率空间**.

例 当仅仅存在可数个可能结果时, 譬如说 $\Omega=\{\omega_1,\omega_2,\cdots\}$, 我们可以取 Ω 的所有子集构成的集合作为 $\mathcal{F}$. 如果 $p_1,p_2,\cdots$ 是非负数且 $\sum_n p_n=1$, 规定

$$P[A]=\sum_{\omega_i\in A}p_i,$$

则 P 是 $\mathcal{F}$ 上的概率测度.

298

把 Ω 的所有子集的集合看作是事件族往往并不是我们所希望的, 也是不合适的. 事实上, 当 Ω 不可数时, 在保证性质 (7.1) 的所有子集的集合上, 定义一个概率测度是不可能的. 无论如何, 我们都规定 $\mathcal{F}$ 使得 (7.1 a) $\sim$ (7.1 c) 成立, 且事件族 $\mathcal{F}$ 应满足

(a) $\varnothing \in \mathcal{F}, \quad \Omega \in \mathcal{F}$;

(b) 若 $A \in \mathcal{F}$, 则$A^c \in \mathcal{F}$, 这里

$$A^c = \{\omega \in \Omega; \omega \notin A\} \text{是 } A \text{ 的余集}; \tag{7.2}$$

(c) $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{F}$, 当 $A_n \in \mathcal{F}, n = 1, 2, \cdots$.

集合 Ω 的子集族 $\mathcal{F}$ 若满足 (7.2 a) $\sim$ (7.2 c) 便称为 σ 域. 如果 $\mathcal{F}$ 是 σ 域, 则

$$\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c\right)^c \in \mathcal{F},$$

只要 $A_n \in \mathcal{F}, n = 1, 2, \cdots$. 显然, $\mathcal{F}$ 中集合的有限并和有限交也属于 $\mathcal{F}$.

在这个结构下, 一个实值随机变量 X 是定义在 Ω 上满足由下面给出的某种 "可测性" 条件的实值函数. 随机变量 X 的分布函数可由下式给出:

$$\Pr\{a < X \leqslant b\} = P[\{\omega : a < X(\omega) \leqslant b\}]. \tag{7.3}$$

换句话说, 随机变量 X 取值在 $(a, b]$ 上的概率就是满足 $a < X(\omega) \leqslant b$ 的结果 ω 集合的概率. 若要 (7.3) 有意义, 则 X 不能是 Ω 上任意函数, 而必须满足条件

$$\{\omega : a < X(\omega) \leqslant b\} \in \mathcal{F}, \qquad \text{对于所有的实数} a < b.$$

因为 $\mathcal{F}$ 正是那些使得 $P[A]$ 有定义的集合 A 的全体. 实际上, 利用 σ 域 $\mathcal{F}$ 的性质 (7.2 a) $\sim$ (7.2 c), 只需要求对于所有的 x,

$$\{\omega : X(\omega) \leqslant x\} \in \mathcal{F}$$

就足够了. 设 $\mathcal{A}$ 是 Ω 子集的任一 σ 域, 我们说 X 是关于 $\mathcal{A}$ 可测的, 或简单地, 是 $\mathcal{A}$ 可测的, 如果对于所有的实值 x

$$\{\omega : X(\omega) \leqslant x\} \in \mathcal{A}.$$

这样, 每个实值随机变量都是 $\mathcal{F}$ 可测的. 一般地说, X 可能关于比较小的 σ 域也是可测的.

299

由随机变量 X 生成的 σ 域定义为使得 X 可测的最小 σ 域, 记为 $\mathcal{F}(X)$. 它是由这样的集合 A 构成的, A 属于任何使得 X 可测的 σ 域. 例如, 如果 X 仅有可数多可能值 $x_1, x_2, \cdots$, 集合

$$A_i = \{\omega : X(\omega) = x_i\}, \quad i = 1, 2, \cdots$$

构成 Ω 的可数分割, 即

$$\Omega = \bigcup_{i=1}^{\infty} A_i,$$

并且

$$A_i \bigcap A_j = \varnothing, \text{ 如果 } i \neq j,$$

这时, $\mathcal{F}(X)$ 是由 $\varnothing, \Omega$ 和 A_i 的所有并集组成.

例 对于那些完全不熟悉这种框架的读者, 下面简单的例子有助于理解这些概念. 我们做如下试验, 丢掷一枚镍币和一枚银币, 并观察其出现正面或反面, 取 Ω 为

$$\Omega = \{(\mathrm{H}, \mathrm{H}), (\mathrm{H}, \mathrm{T}), (\mathrm{T}, \mathrm{H}), (\mathrm{T}, \mathrm{T})\},$$

其中, 举例说, (H, T) 表示结果 "镍币是正面, 银币是反面". 我们取 Ω 的所有子集的集合作为事件族. 在 Ω 中假设每个结果是等可能的, 我们得到概率测度为下表所示:

$A \in \mathcal{F}$	$P[A]$	$A \in \mathcal{F}$	$P(A)$
$\varnothing$	0	Ω	1
{(H, H)}	$\frac{1}{4}$	{(H, T), (T, H), (T, T)}	$\frac{3}{4}$
{(H, T)}	$\frac{1}{4}$	{(H, H), (T, H), (T, T)}	$\frac{3}{4}$
{(T, H)}	$\frac{1}{4}$	{(H, H), (H, T), (T, T)}	$\frac{3}{4}$
{(T, T)}	$\frac{1}{4}$	{(H, H), (H, T), (T, H)}	$\frac{3}{4}$
{(H, H), (H, T)}	$\frac{1}{2}$	{(T, H), (T, T)}	$\frac{1}{2}$
{(H, H), (T, H)}	$\frac{1}{2}$	{(H, T), (T, T)}	$\frac{1}{2}$
{(H, H), (T, T)}	$\frac{1}{2}$	{(H, T), (T, H)}	$\frac{1}{2}$

事件 "镍币是正面" 为 {(H, H), (H, T)}, 按照上表其概率是 $\frac{1}{2}$, 这与实际相符. 300

若镍币是正面, 令 X_n 为 1; 若镍币是反面, 令 X_n 为 0. 类似地, X_d 是对应于银币的随机变量. 令 $Z = X_n + X_d$ 是正面总数. 作为 Ω 的函数, 我们有

$\omega \in \Omega$	$X_n(\omega)$	$X_d(\omega)$	$Z(\omega)$
(H ,H)	1	1	2
(H, T)	1	0	1
(T, H)	0	1	1
(T, T)	0	0	0

最后, 由 X_n 和 Z 产生的 σ 域是

$$\mathcal{F}(X_n)=\varnothing,\Omega,\{(\mathrm{H},\mathrm{H}),(\mathrm{H},\mathrm{T})\},\{(\mathrm{T},\mathrm{H}),(\mathrm{T},\mathrm{T})\},$$

和

$$\mathcal{F}(Z)=\varnothing,\Omega,\{(\mathrm{H},\mathrm{H})\},\{(\mathrm{H},\mathrm{T}),(\mathrm{T},\mathrm{H})\},\{(\mathrm{T},\mathrm{T})\},$$

$$\{(\mathrm{H},\mathrm{T}),(\mathrm{T},\mathrm{H}),(\mathrm{T},\mathrm{T})\},\{(\mathrm{H},\mathrm{H}),(\mathrm{T},\mathrm{T})\},$$

$$\{(\mathrm{H},\mathrm{H}),(\mathrm{H},\mathrm{T}),(\mathrm{T},\mathrm{H})\}.$$

$\mathcal{F}(X_n)$ 包含 4 个集合, $\mathcal{F}(Z)$ 包含 8 个集合. 请读者考虑, X_n 关于 $\mathcal{F}(Z)$ 是否可测? 或 Z 是否关于 $\mathcal{F}(X_n)$ 可测?

每一对随机变量 X,Y 确定一个 σ 域称为由 X,Y 生成的 σ 域. 它是使得 X 和 Y 均可测的最小 σ 域. 这个 σ 域是由这样的集合 A 组成的, A 属于任何使得 X 和 Y 都可测的 σ 域. 如果 X 和 Y 假定都仅有可数多个可能值, 譬如说, 分别为 $x_1,x_2,\cdots$ 和 $y_1,y_2,\cdots$, 则集合

$$A_{ij}=\{\omega:X(\omega)=x_i,Y(\omega)=y_j\},\quad i,j=1,2,\cdots,$$

是 Ω 的一个分割, 并且 $\mathcal{F}(X,Y)$ 正好是由 $\varnothing,\Omega$ 及 A_{ij} 的所有并集组成的. 注意到 X 关于 $\mathcal{F}(X,Y)$ 可测, 故 $\mathcal{F}(X)\subset\mathcal{F}(X,Y)$.

更加一般地, 令 $\{X(t);t\in T\}$ 是任一随机变量族, 则 $\{X(t);t\in T\}$ 所产生的 σ 域是使得每个随机变量 $X(t),t\in T$ 都可测的最小 σ 域, 记为 $\mathcal{F}\{X(t);t\in T\}$.

由实数集合构成的一个 σ 域起着特别作用. 这就是著名的 Borel 集的 σ 域, 它是由函数 $f(x)=x,x\in(-\infty,\infty)$ 产生的 σ 域, 它也是包含一切区间 $(a,b],-\infty\leqslant a\leqslant b<+\infty$ 的最小 σ 域. 一个实变量的实值函数如果关于 Borel 集的 σ 域是可测的, 称为 Borel 可测的.

301

在 n 维欧几里得空间, Borel 集的 σ 域是由函数族

$$\{f_i(x_1,\cdots,x_n)=x_i;i=1,2,\cdots,n\}$$

产生的 σ 域.

关于 σ 域的条件期望

尽管每个随机变量 Y 都产生一个 σ 域 $\mathcal{F}(Y)$, 但并不是说每个 σ 域都是由这种形式生成的. 因此下面关于 σ 域的条件期望是关于随机变量的条件期望的推广.

我们先从其严格的定义开始.

定义 7.1 设 X 是概率空间 $(\Omega, \mathcal{F}, P)$ 上的一个随机变量, 且 $E[\,|X|\,] < \infty$. 令 $\mathcal{B}$ 是 $\mathcal{F}$ 中的一个 σ 域, 即每个集合 $B \in \mathcal{B}$ 都是 $\mathcal{F}$ 的元素. 关于 $\mathcal{B}$ 的条件期望定义为具有如下性质的随机变量 $E[\,X|\mathcal{B}\,]$:

(i) $E[X|\mathcal{B}]$ 是关于 $\mathcal{B}$ 的可测函数;

(ii) 对于所有的 $B \in \mathcal{B}, E[XI_B] = E\{E[X|\mathcal{B}]I_B\}$, 其中 I_B 是 B 的示性函数.

条件 (ii) 等价于条件

(ii′) 对每个有界的 $\mathcal{B}$ 可测随机变量 $Z, E[XZ] = E\{E[X|\mathcal{B}]Z\}$.

关于上述定义有如下几点说明. 首先, 可以证明当 $E[|X|] < \infty$ 时, 满足 (i) 和 (ii) 的随机变量是存在的, 所以定义是非空的. 事实上, 只要 $E[X^+]$ 和 $E[X^-]$ 不同时为无限, 上述定义即有意义. 另一方面, 满足定义中 (i)(ii) 的条件期望 $E[X|\mathcal{B}]$ 可能不止一个. 这样, 我们可以谈及条件期望的不同"变形". 幸运的是, 任何两个变形以概率 1 都是相等的. 即, 如果 $E^{(1)}[X|\mathcal{B}]$ 和 $E^{(2)}[X|\mathcal{B}]$ 满足 (i) 和 (ii), 则

$$P\{E^{(1)}[X|\mathcal{B}] = E^{(2)}[X|\mathcal{B}]\} = 1.$$

因此, 从概率意义上看, 条件期望定义中不确定性不会引起麻烦. 最后让我们说明 (ii) 和 (ii′) 的等价性. 显然 (ii′) 可以推出 (ii), 因为我们可取 Z 是如下有界的 $\mathcal{B}$ 可测函数

$$Z(\omega) = I_B(\omega) = \begin{cases} 1, & 若\ \omega \in B, \\ 0, & 若\ \omega \notin B. \end{cases}$$

302

相反的蕴涵关系是由于任何可积函数可以通过适当的阶梯函数

$$Z_n(\omega) = \sum_{i=1}^{n} \alpha_{in} I_{B_i}(\omega)$$

来逼近, 其中 $\{B_1, \cdots, B_n\}$ 是 Ω 的有限分割且 $B_i \in \mathcal{B}$. 这样, (ii′) 对 Z_n 是成立的, 通过极限过渡, (ii′) 对任何有界 $\mathcal{B}$ 可测随机变量也成立.

在第 1 章我们将已知 $Y = y$ 时 X 的条件期望表达为

(I) $E[X|Y = y]$ 是 y 的函数, 且满足

(II) 对每个有界函数 g,

$$E[Xg(Y)] = \int E[X|Y = y]g(y)\mathrm{d}F_Y(y).$$

我们希望现在的条件期望的定义可以作为前面定义的推广, 即当 $\mathcal{B} = \mathcal{F}(Y)$ 是由随机变量 Y 产生的 σ 域时与前面的定义相吻合. 可以证明随机变量 Z 关于 $\mathcal{B} = \mathcal{F}(Y)$ 可测当且仅当 $Z = f(Y)$, 其中 f 是某个 Borel 可测函数. (问题 13 要求在 $\mathcal{B}$ 是某可列分割 $\mathcal{B}_0$ 的并的情况下证明这个命题.) 这样, (I) 说明 $E[X|Y]$ 是关

于 $\mathcal{B} = \mathcal{F}(Y)$ 可测的, 而 (II) 可推得 (ii′), 即对于所有有界的 $Z = f(Y), E[XZ] = E\{E[X|Y]Z\}$ 成立. 因此, 定义 7.1 推广了关于随机变量的条件期望.

最后, 我们来讨论一个特殊情形, $\Omega = \{\omega_1, \omega_2, \cdots\}$ 是可数的, $\mathcal{F}$ 由 Ω 的所有子集构成, P 由下式给出

$$P[A] = \sum_{\omega_i \in A} p_i, \qquad A \in \mathcal{F}, \tag{7.4}$$

其中 $p_1, p_2, \cdots$ 是和为 1 的非负数. 随机变量 X 的期望定义为

$$E[X] = \sum_{j=1}^{\infty} X(\omega_j) p_j,$$

假定上述级数是绝对收敛的. 令 $\mathcal{B}_0 = \{B_1, B_2, \cdots\}$ 是 Ω 的可数分割, 并令 $\mathcal{B}$ 是由 ϕ, Ω 和 $\mathcal{B}_0$ 中所有可能的并集构成的 σ 域. 对于 $\mathcal{B}_0$ 中每个 B_j, 定义初等条件期望为

$$E[X|B_j] = \sum_{\omega_k \in B_j} X(\omega_k) P[\{\omega_k\}|B_j],$$

其中

303

$$P[A|B_j] = P[A \bigcap B_j]/P[B_j], \qquad A \in \mathcal{F}. \tag{7.5}$$

在 $P[B_j] = 0$ 的情况下定义是无效的, 因为这时式子右边出现 $\dfrac{0}{0}$. 当 $P[B_j] = 0$ 时任意置 $E[X|B_j] = 17$, 这样就完成了定义. 当然这里也出现了如定义 7.1 附注中所指出不唯一的问题. 下一步是按如下公式定义随机变量 $E[X|\mathcal{B}]$

$$\begin{aligned} E[X|\mathcal{B}](\omega) &= E[X|B_j], \qquad \text{如果} \omega \in B_j, \\ &= \sum_{j=1}^{\infty} E[X|B_j] I_{B_j}(\omega), \end{aligned}$$

其中

$$I_{B_j}(\omega) = \begin{cases} 1, & \text{若 } \omega \in B_j, \\ 0, & \text{若 } \omega \notin B_j. \end{cases}$$

这样, 随机变量 $E[X|\mathcal{B}]$ 在每个集合 B_j 上是一常数, 因此是 $\mathcal{B}$ 可测的. 我们检验定义 7.1 中的 (ii). 令 $B \in \mathcal{B}$ 是任意指定的, 譬如说 $B = \bigcup_{n=1}^{\infty} B'_n$, 其中 $B'_n \in \mathcal{B}_0, n = 1, 2, \cdots$. 一方面我们有

$$E[XI_B] = \sum_{n=1}^{\infty} \sum_{\omega_k \in B'_n} X(\omega_k) p_k,$$

而另一方面我们有

$$E\{E[X|\mathcal{B}]I_B\}=\sum_{n=1}^{\infty}E[X|B_n']P[B_n']=\sum_{n=1}^{\infty}\sum_{\omega_k\in B_n'}X(\omega_k)p_k.$$

这就验证了 (ii).

现在我们说明, 当 $P[B_j]>0$ 时 $E[X|B_j]$ 的初等定义可以由定义 7.1 中取 $B=B_j$ 得到. 这时 (ii) 变成

$$\sum_{\omega_k\in B_j}X(\omega_k)p_k=\sum_{\omega_k\in B_j}E[X|\mathcal{B}](\omega_k)p_k. \tag{7.6}$$

$E[X|\mathcal{B}]$ 是 $\mathcal{B}$ 可测的当且仅当它在每个 B_j 上取常数值, 记为 a_j. 这样, (7.6) 变为

$$\sum_{\omega_k\in B_j}X(\omega_k)p_k=a_jP[B_j],$$

并且常数值 a_j 一定是

$$\begin{aligned}a_j&=\sum_{\omega_k\in B_j}X(\omega_k)p_k/P[B_j]\\&=E[X|B_j],\end{aligned}$$

只要 $P[B_j]>0$. 304

这样一来, 当 $\mathcal{B}$ 是可数分割 $\mathcal{B}_0$ 的并的 σ域时, 定义 7.1 包含了条件期望的直观定义.

关于随机变量的条件期望是在条件分布之下计算的期望. 当我们把 "相等" 理解为 "以概率 1 相等", 则关于 σ 域的条件期望保留了关于随机变量的条件期望大部分熟知的性质. 为了便于参考, 我们列出这性质, 它们对应于第 1 章的 (1.6)、(1.7) 和 (1.10)~(1.14). 但关于 (1.8), 没有比较满意的类似式子. 我们假设 X_1,X_2 和 X 是具有限期望的随机变量, a_1,a_2 为实数, $\mathcal{A}$ 和 $\mathcal{B}$ 是 $\mathcal{F}$ 的子 σ 域. 则

$$E[a_1X_1+a_2X_2|\mathcal{B}]=a_1E[X_1|\mathcal{B}]+a_2E[X_2|\mathcal{B}]; \tag{7.7}$$

$$X\geqslant 0 \text{ 蕴涵着 } E[X|\mathcal{B}]\geqslant 0; \tag{7.8}$$

$$E[XZ|\mathcal{B}]=Z\cdot E[X|\mathcal{B}], \text{对每个有界 }\mathcal{B}\text{可测函数}Z; \tag{7.9}$$

$$E[XZ]=E\{Z\cdot E[X|\mathcal{B}]\}, \text{对每个有界}\mathcal{B}\text{可测函数}Z; \tag{7.10}$$

$$E[Z|\mathcal{B}]=Z, \text{ 若 } Z \text{ 是 }\mathcal{B}\text{可测且满足 } E[\,|Z|\,]<\infty; \tag{7.11}$$

$$E[X|\mathcal{A}] = E\{E[X|\mathcal{B}]|\mathcal{A}\}, \text{如果 } \mathcal{A} \subset \mathcal{B}; \tag{7.12}$$

$$E[X] = E\{E[X|\mathcal{B}]\} \text{ (全概率公式)}. \tag{7.13}$$

验证这些式子的方法都是类似的, 我们验证 (7.7) 作为例子. 我们证明 $a_1E[X_1|\mathcal{B}] + a_2E[X_2|\mathcal{B}]$ 满足 $E[a_1X_1 + a_2X_2|\mathcal{B}]$ 的定义条件. 首先, 它是两个 $\mathcal{B}$可测随机变量的线性组合, 也是 $\mathcal{B}$可测的. (问题 14 要求当 $\mathcal{B}$ 是可数分割 $\mathcal{B}_0$ 的并的 σ域时证明这个命题.) 这样, $a_1E[X_1|\mathcal{B}]+a_2E[X_2|\mathcal{B}]$ 满足定义 7.1 的 (i). 然后用 $a_1E[X_1|\mathcal{B}]+a_2E[X_2|\mathcal{B}]$ 代替 $E[a_1X_1 + a_2X_2|\mathcal{B}]$ 检验 (ii′) 是否满足. 考虑任何一个有界 $\mathcal{B}$ 可测函数 Z, 我们有

$$\begin{aligned} E[\{a_1X_1 + a_2X_2\}Z] &= a_1E[X_1Z] + a_2E[X_2Z] \\ &= a_1E\{E[X_1|\mathcal{B}]Z\} + a_2E\{E[X_2|\mathcal{B}]Z\} \\ &= E\{(a_1E[X_1|\mathcal{B}] + a_2E[X_2|\mathcal{B}])Z\}. \end{aligned}$$

305 这就完成了 (7.7) 的验证.

关于 σ 域递增族的鞅

设 $\{X_n\}_0^\infty$ 是概率空间 $(\Omega, \mathcal{F}, P)$ 上的一实随机变量序列. 令 $\{\mathcal{F}_n\}$ 是 $\mathcal{F}$ 的一列子 σ 域, 满足

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n \subset \cdots \subset \mathcal{F}.$$

如果对每个 n, X_n 是 $\mathcal{F}_n$可测的, 则称 $\{X_n\}$**适应于**$\{\mathcal{F}_n\}$. 例如, 假设 $Y_0, Y_1, \cdots$ 也定义在 $(\Omega, \mathcal{F}, P)$ 上, 并且 $\mathcal{F}_n$ 是由 $\{Y_0, Y_1, \cdots, Y_n\}$ 产生的 σ 域. 则对于所有的 n

$$\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F},$$

并且如果 $\{g_n(\cdot, \cdots, \cdot)\}$ 是一列 Borel 可测函数而 $X_n = g_n(Y_0, \cdots, Y_n)$. 则 $\{X_n\}$ 适应于 $\{\mathcal{F}_n\}$.

正像早先对于 $(Y_0, \cdots, Y_n)$ 所叙述的那样. 我们再次把 $\mathcal{F}_n$ 看成包含时刻 n 为止得到的信息, 则 X_n 关于 $\mathcal{F}_n = \mathcal{F}(Y_0, \cdots, Y_n)$ 可测当且仅当 X_n 可由 $(Y_0, \cdots, Y_n)$ 所确定.

关系 $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ 表示随着 n 增加, 信息不断增加.

定义 7.2 设 $\{X_n\}$ 是定义在概率空间 $(\Omega, \mathcal{F}, P)$ 上的一列随机变量. 令 $\{\mathcal{F}_n\}$ 是 $\mathcal{F}$ 的一列子 σ 域且对于所有的 n 有

$$\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}.$$

$\{X_n\}$ 称为关于 $\{\mathcal{F}_n\}$ 的下鞅, 如果:

(i) $\{X_n\}$ 适应于 $\{\mathcal{F}_n\}$(即每个 X_n 是 $\mathcal{F}_n$ 可测的),

(ii) 对于所有的 n, $E[X_n^+] < \infty$,

(iii) 对于所有的 n, $E[X_{n+1}|\mathcal{F}_n] \geqslant X_n$.

如果 $\{-X_n\}$ 是下鞅, 则 $\{X_n\}$ 称为上鞅. 如果 $\{X_n\}$ 和 $\{-X_n\}$ 皆为下鞅, 则 $\{X_n\}$ 称为关于 $\{\mathcal{F}_n\}$ 的鞅.

下面的说明可能对读者有所帮助. 如果 $\{X_n\}$ 是鞅, 则

$$X_n = E[X_{n+1}|\mathcal{F}_n]. \tag{7.14}$$

由条件期望的定义, 右边部分是 $\mathcal{F}_n$可测的, 所以右边也是 $\mathcal{F}_n$可测的. 因此, 对于鞅由 (7.14) 可推得 (i). 利用条件期望的性质, 条件 (iii) 可等价地叙述为

(iii′) 对于所有的有界 $\mathcal{F}_n$ 可测函数 $Z \geqslant 0$,

$$E[X_{n+1}Z] \geqslant E[X_nZ].$$

306

为了由 (iii) 推得 (iii′), 我们利用关于 σ 域条件期望的定义和性质 (7.7)、(7.8) 得到

$$E[X_{n+1}Z] = E\{E[X_{n+1}|\mathcal{F}_n]Z\} \geqslant E[X_nZ].$$

为了从 (iii) 导出 (iii′), 利用

$$E\{E[X_{n+1}|\mathcal{F}_n]Z\} = E[X_{n+1}Z]$$

得到对于所有的有界 $\mathcal{F}_n$ 可测函数 $Z \geqslant 0$,

$$E\{(E[X_{n+1}|\mathcal{F}_n] - X_n)Z\} \geqslant 0.$$

现在 $Y = E[X_{n+1}|\mathcal{F}_n] - X_n$ 是 $\mathcal{F}_n$ 可测的, 并且如果对于所有的有界 $\mathcal{F}_n$ 可测函数 $Z \geqslant 0$ 有 $E[YZ] \geqslant 0$, 则 $Y \geqslant 0$.(问题 15 要求在 $\mathcal{B}$ 是由可数分割 B_0 的并集构成的 σ 域的情况下证明此命题.) 由此,

$$Y = E[X_{n+1}|\mathcal{F}_n] - X_n \geqslant 0.$$

这就是 (iii).

本章开头叙述的所有关于随机变量的鞅的结果均可以推广到关于增 σ 域的鞅上, 只需在形式上作些改变. 例如, 我们一旦引进关于 σ 域族的马尔可夫时间的定义, 可选抽样定理以及鞅的收敛定理都仍然成立. 我们将不再重述全部结果, 仅叙述关于鞅的一些结果, 这主要在教学上有助于我们熟练地运用关于 σ 域的条件期望性质.

命题 7.1 设 $\{X_n\}$ 是关于 $\{\mathcal{F}_n\}$ 的鞅. 则对于所有的 n, $E[X_n] = E[X_0]$.

证明 应用全概率公式 (7.13) 于鞅等式 $X_n = E[X_{n+1}|\mathcal{F}_n]$, 我们得到

$$E[X_n] = E\{E[X_{n+1}|\mathcal{F}_n]\} = E[X_{n+1}].$$

使用归纳法即可完成证明. ■

命题 7.2 设 $\{X_n\}$ 是关于 $\{\mathcal{F}_n\}$ 的鞅, 如果 Z 是有界 $\mathcal{F}_n$可测随机变量, 则

307

$$E[ZX_{n+k}|\mathcal{F}_n] = ZX_n, \quad n = 0, 1, \cdots; k \geqslant 1.$$

证明 我们使用性质 (7.9), 对于有界 $\mathcal{F}_n$可测的 Z, 有

$$E[ZX_{n+k}|\mathcal{F}_n] = ZE[X_{n+k}|\mathcal{F}_n]. \tag{7.15}$$

再根据 (7.12), 得

$$E[X_{n+k}|\mathcal{F}_n] = E\{E[X_{n+k}|\mathcal{F}_{n+k-1}]|\mathcal{F}_n\} = E[X_{n+k-1}|\mathcal{F}_n],$$

利用归纳法易得

$$E[X_{n+k}|\mathcal{F}_n] = E[X_{n+1}|\mathcal{F}_n] = X_n,$$

由此结果与 (7.15) 即得结论. ■

设 $\{\mathcal{F}_n\}$ 是满足 $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$(对于所有的 n) 的一列子 σ域, 取值在 $\{0, 1, \cdots, \infty\}$ 上的随机变量 T 称为关于 $\{\mathcal{F}_n\}$ 的马尔可夫时间, 如果对于每个 $n = 0, 1, 2, \cdots$, 事件 $\{T = n\}$ 属于 $\mathcal{F}_n$. 由于每个随机变量都是定义在样本空间 Ω 上的函数, 这样, 我们要求

$$\{\omega : T(\omega) = n\} \in \mathcal{F}_n, \qquad \text{对于所有的} n. \tag{7.16}$$

由于每个 $\mathcal{F}_n$ 是一个 σ域并且对于所有的 n 有 $\mathcal{F}_n \subset \mathcal{F}_{n+1}$. 式 (7.16) 等价于

$$\{\omega : T(\omega) \leqslant n\} \in \mathcal{F}_n, \qquad \text{对于所有的} n, \tag{7.17}$$

或

$$\{\omega : T(\omega) > n\} \in \mathcal{F}_n, \qquad \text{对于所有的} n. \tag{7.18}$$

例如, 为从 (7.16) 推得 (7.17), 注意

$$\{\omega : T(\omega) \leqslant n\} = \bigcup_{k=0}^{n} \{\omega : T(\omega) = k\}.$$

由于每个 $\{\omega : T(\omega) = k\} \in \mathcal{F}_k \subset \mathcal{F}_n$, 所以其并集属于 $\mathcal{F}_n$, 因此满足 (7.17).

等价地, 我们也可要求每个示性函数

$$I_{\{T=n\}}=\begin{cases}1, & \text{如果 } T=n,\\ 0, & \text{如果 } T\neq n,\end{cases}$$

是关于 $\mathcal{F}_n$ 可测的.

308

如果 $\mathcal{F}_n=\mathcal{F}(Y_0,Y_1,\cdots,Y_n)$ 是由随机变量组 $Y_0,Y_1,\cdots,Y_n$ 产生的 σ 域, 则 $I_{\{T=n\}}$ 关于 $\mathcal{F}_n$ 可测当且仅当

$$I_{\{T=n\}}=g_n(Y_0,\cdots,Y_n),$$

其中 $g_n(\cdot,\cdots,\cdot)$ 是某个适当的可测函数. 这样, 后面马尔可夫时间的定义是早先定义的推广.

每个常数 $T\equiv n$ 是马尔可夫时间. 如果 T 和 S 是马尔可夫时间, $T+S, T\wedge S=\min\{T,S\}$ 和 $T\vee S=\max\{T,S\}$ 也是马尔可夫时间. 例如, 为证明 $T\wedge S$ 是马尔可夫时间, 我们有

$$\{\omega:T\wedge S>n\}=\{\omega:T>n\}\bigcap\{\omega:S>n\}\in\mathcal{F}_n.$$

引理 7.1 设 $\{X_n\}$ 是鞅, T 是关于 $\{\mathcal{F}_n\}$ 的马尔可夫时间. 则对于所有的 $n\geqslant k$, 有

$$E[X_nI_{\{T=k\}}]=E[X_kI_{\{T=k\}}].$$

($I_{\{\cdot\}}$ 是由 $\{\cdot\}$ 所描述事件的示性函数.)

证明 我们利用事实: $I_{\{T=k\}}$ 是 $\mathcal{F}_k$可测的. 则

$$\begin{aligned}E[X_nI_{\{T=k\}}]&=E\{E[X_nI_{\{T=k\}}|\mathcal{F}_k]\} && (\text{由大数定律})\\ &=E\{I_{\{T=k\}}E[X_n|\mathcal{F}_k]\} && (\text{由}(7.10))\\ &=E[X_kI_{\{T=k\}}].\end{aligned}$$

■

鉴于上述. 关于 σ 域族的鞅的研究完全平行于 6.1~6.6 节所进行的讨论. 例如, 我们有下面引理:

引理 7.2 设 $\{X_n\}$ 是鞅. T 是关于 $\{\mathcal{F}_n\}$ 的马尔可夫时间. 则对于所有的 $n=0,1,\cdots$,

$$E[X_0]=E[X_{T\wedge n}]=E[X_n].$$

证明 如在引理 3.2 所做的那样, 根据引理 7.1, 在细节上做些适当修改即可得证. ■

关于上穿不等式、最大值不等式、可选抽样定理以及鞅的收敛定理, 均可由引理 7.2 推得, 正如我们前面从引理 3.2 所导出的一样, 下面我们举一个例子.

例 设 $\{Y_n\}$ 是定义在某概率空间 $(\Omega, \mathcal{F}, P)$ 上的随机变量序列, 并且对于每个 n, 令 $\mathcal{F}_n$ 是由 $(Y_0, \cdots, Y_n)$ 产生的 σ域. 如果 Z 满足 $E[\,|Z|\,] < \infty$, 我们在 6.1 节例 (k) 中指出了

309

$$X_n = E[Z|\mathcal{F}_n], \qquad n = 0, 1, \cdots,$$

构成一个鞅, 并且在 6.6 节例 (f) 我们验知这是一致可积鞅. 由鞅收敛定理知极限

$$\lim_{n\to\infty} X_n = X_\infty$$

以概率 1 存在, 并且成立

$$\lim_{n\to\infty} E[\,|X_n - X_\infty|\,] = 0.$$

我们提到过 X_∞ 可以表示为

$$X_\infty = E[Z|Y_0, Y_1, \cdots].$$

若把上式右边部分解释为 Z 在条件分布下的期望则是比较困难的. 然而, 如果我们用 $\mathcal{F}_\infty$ 表示由 $(Y_0, Y_1, \cdots)$ 产生的 σ域, 我们不难得到

$$X_\infty = E[Z|\mathcal{F}_\infty].$$

依照定义 7.1, 我们首先需要证明 X_∞ 是 $\mathcal{F}_\infty$可测的, 然后对每个有界 $\mathcal{F}_\infty$可测随机变量 W, 证明等式

$$E[X_\infty W] = E[ZW]. \tag{7.19}$$

每个 X_n 是 $\mathcal{F}_n$可测的, 因此是 $\mathcal{F}_\infty$可测的, 因为对于所有的 $n, \mathcal{F}_n \subset \mathcal{F}_\infty$. 由此推得 $X_\infty = \lim_{n\to\infty} X_n$ 是 $\mathcal{F}_\infty$可测的. 从另外一个观点看, 由于每个 $X_n = E[Z|Y_0, \cdots, Y_n]$ 是 $Y_0, \cdots, Y_n$ 的函数, 所以 $X_\infty = \lim_{n\to\infty} X_n$ 是整个序列 $Y_0, Y_1, \cdots$ 的一个适当可测函数, 因此关于 $\mathcal{F}_\infty$ 可测.

为证明 (7.19), 只要考虑任一有界 $\mathcal{F}_m$可测函数 W_m 即可, 其中 m 是任意正整数. 让 m 增加, 用随机变量 W_m 逼近 $\mathcal{F}_\infty$可测函数 W, 便可推得一般情形. 但是, 当 W_m 是 $\mathcal{F}_m$可测时,

$$\begin{aligned} E[X_n W_m] &= E\{E[Z|\mathcal{F}_n]W_m\} \\ &= E\{E[ZW_m|\mathcal{F}_n]\}, \qquad \text{如果 } n \geqslant m \\ &\qquad (\text{因为} W_m \text{是} \mathcal{F}_m \subset \mathcal{F}_n \text{可测的}) \\ &= E[ZW_m], \qquad (\text{由全概率公式}) \end{aligned}$$

关于 n 取极限, 得

$$E[X_\infty W_m] = \lim_{n\to\infty} E[X_n W_m] = E[ZW_m],$$

因此 (7.19) 得证.

310

我们先来说明另外一个重要事实. 由于 $X_\infty = \lim_{n\to\infty} X_n$ 是 $\mathcal{F}_\infty$可测的, 我们从 (7.11) 知道

$$X_\infty = E[X_\infty|\mathcal{F}_\infty].$$

而另一方面, 我们刚刚验证了表达式

$$X_\infty = E[Z|\mathcal{F}_\infty],$$

根据性质 (7.12) 和 $\mathcal{F}_n \subset \mathcal{F}_\infty$, 我们有

$$\begin{aligned} X_n &= E[Z|\mathcal{F}_n] \\ &= E\{E[Z|\mathcal{F}_\infty]|\mathcal{F}_n\} \\ &= E[X_\infty|\mathcal{F}_n]. \end{aligned}$$

即, $X_n = E[X_\infty|\mathcal{F}_n]$, 其中 $X_\infty = \lim_{n\to\infty} X_n$. 这个事实对于每个一致可积鞅都成立, 我们写成如下的引理.

引理 7.3 设 $\{X_n\}$ 是关于 $\{\mathcal{F}_n\}$ 的一致可积鞅 (见 6.3 节), 则

$$X_n = E[X_\infty|\mathcal{F}_n],$$

其中

$$X_\infty = \lim_{n\to\infty} X_n.$$

证明 鞅收敛定理保证极限 $X_\infty = \lim_{n\to\infty} X_n$ 以概率 1 存在并且成立

$$\lim_{n\to\infty} E[\,|X_n - X_\infty|\,] = 0. \tag{7.20}$$

现在我们证明 X_n 具有 $E[X_\infty|\mathcal{F}_n]$ 按定义 7.1 所要求的三个性质. 首先, 由于 $\{X_n\}$ 是鞅, 因此 X_n 是 $\mathcal{F}_n$ 可测的. 令 W 是有界 $\mathcal{F}_n$ 可测随机变量, 则

$$\begin{aligned} E[X_\infty W] &= E[\lim_{m\to\infty} X_m W] \\ &= \lim_{m\to\infty} E[X_m W] \qquad \text{(极限与期望交换的验证在下面)} \\ &= E[X_n W]. \end{aligned}$$

311

即 $X_n = E[X_\infty|\mathcal{F}_n]$. 极限号交换的合理性依据是 (7.20) 和不等式

$$\begin{aligned}|E[X_\infty W] - E[X_n W]| &\leqslant E[\,|X_\infty W - X_n W|] \\ &\leqslant AE[\,|X_n - X_\infty|\,],\end{aligned}$$

其中 $|W| \leqslant A, A < \infty$. 引理证毕. ■

我们已经建立了重要结果, 即每个一致可积鞅 $\{X_n\}$ 具有 Doob 过程 $X_n = E[Z|\mathcal{F}_n]$ 的形式, 其中 $Z = X_\infty = \lim\limits_{n\to\infty} X_n$.

应用于数学分析

设 f 是在 $[0,1]$ 上实值函数, 且是利普希茨连续的. 即 f 满足

$$|f(x) - f(y)| \leqslant C|x-y|, \qquad \text{对于所有的} x, y \in [0,1],$$

其中 $C < \infty$ 是常数. 对于 $n = 1, 2, \cdots$ 记 $\mathcal{P}_n$ 是由下面给出的 $[0,1)$ 的分割:

$$\mathcal{P}_n = \{[k/2^n, [k+1]/2^n);\ k = 0, 1, \cdots, 2^n - 1\}.$$

而 $\mathcal{F}_n$ 是由 $\phi, \Omega = [0,1)$ 和 $\mathcal{P}_n$ 中集合的并集构成的 σ 域.

设 Z 在 $[0,1)$ 上具有均匀分布, 并定义随机变量序列

$$X_n = 2^n\{f(k/2^n) - f((k-1)/2^n)\},$$

$$\text{如果}(k-1)/2^n \leqslant Z < k/2^n.$$

则 $\{X_n\}$ 是关于 $\{\mathcal{F}_n\}$ 的鞅. 事实上, $\mathcal{F}_n$ 是由 $Y_0, Y_1, \cdots, Y_n$ 产生的 σ 域, 其中

$$Y_n = k/2^n, \text{当} k/2^n \leqslant Z < (k+1)/2^n,$$

我们在 6.1 节例 (l) 中已验证了它是一个鞅.

注意到 X_n 近似地是 f 在 (随机选择的) 点 Z 上的导数. 当然, f 可以是不可微的, 但却是利普希茨连续的, 对于所有的 $n, |X_n| \leqslant C$, 因此 $\{X_n\}$ 是一致可积的, 故

$$X_\infty = \lim_{n\to\infty} X_n$$

对于 $Z \in [0,1)$ 的具有概率 1 的集合存在. 由引理 7.3, 我们有

$$X_n = E[X_\infty|\mathcal{F}_n]. \tag{7.21}$$

312 显然, X_∞ 是随机变量 Z 的某个函数 g, 记为 $X_\infty = g(Z)$.

取 $B=[0,k/2^n)\in\mathcal{F}_n$, 则由 (7.21) 有

$$\begin{aligned}E[X_nI_B]&=f(k/2^n)-f(0)\\&=E[X_\infty I_B]\\&=\int_0^{k/2^n}g(x)\mathrm{d}x.\end{aligned}$$

对任意 $z\in[0,1)$ 可用一个二进位有理数序列收敛于它. 在上式中按此序列取极限, 得

$$f(z)-f(0)=\int_0^z g(x)\mathrm{d}x.$$

在这个意义下, g 是 f 的导数, 这就是所谓拉东 - 尼古丁导数.

6.8 其他类型的鞅

鞅的概念仅要求过程 $\{X(t);t\in T\}$ 的指标集 T 具有某种顺序结构. 特别, T 可以是实直线上任意子集.

定义 8.1 设 T 是 $(-\infty,+\infty)$ 中的一个子集, 令 $\{X(t);t\in T\}$ 是定义在概率空间 $(\Omega,\mathcal{F},P)$ 上的随机过程. 对于每个 $t\in T$, 假设 $\mathcal{F}_t$ 是 $\mathcal{F}$ 的子 σ 域并且

$$\mathcal{F}_t\subset\mathcal{F}_s,\qquad \text{如果}t<s,t,s\in T.$$

称 $\{X(t)\}$ 为关于 $\{\mathcal{F}_t\}$ 的下鞅, 如果对于所有的 $t\in T$ 有

(i) $X(t)$ 是 $\mathcal{F}_t$ 可测的,

(ii) $E[X(t)^+]<\infty$, 并且

(iii) $E[X(t+u)|\mathcal{F}_t]\geqslant X(t),\quad u>0,t+u\in T$.

此外, 如果 $\{-X(t)\}$ 是下鞅则称 $\{X(t)\}$ 为上鞅. 若 $\{X(t)\}$ 既是下鞅又是上鞅, 则称为鞅.

指标 T 通常有下面几种情况:

$T=\{\cdots,-2,-1,0\}$, 负整数集合,

$T=\{\cdots,-1,0,1,\cdots\}$, 所有整数集合,

$T=[0,\infty)$, 正实直线,

$T=(-\infty,\infty)$, 实直线,

甚至

$T=\mathbf{Q}$, 有理数集合.

反向鞅

设 $\{X_n; n=0,-1,-2,\cdots\}$ 是关于 $\{\mathcal{F}_n; n=0,-1,-2,\cdots\}$ 的下鞅. 作为具体例子, 可以假设 $\mathcal{F}_n$ 是由随机变量序列 $\{Y_n, Y_{n-1}, Y_{n-2},\cdots\}$ 产生的 σ 域. 当然也可能是其他情况.

引理 5.1 的最大值不等式变为

$$\lambda \Pr\left\{\max_{n\leqslant k\leqslant 0} X_k > \lambda\right\} \leqslant E[X_0], \quad \lambda > 0, \tag{8.1}$$

其中假设每个 $X_n \geqslant 0$, 并且因右边部分与 n 无关, 故

$$\lambda \Pr\left\{\sup_{k\leqslant 0} X_k > \lambda\right\} \leqslant E[X_0], \qquad \lambda > 0.$$

我们发现 6.4 节的上穿不等式 (4.11) 也可作同样的改进. 已知实数 $a<b$ 和负整数 N, 定义 $V_{a,b}(N)$ 是满足下列条件的数对 (i,j) 的数目：$N\leqslant i<j\leqslant 0$, 并且不等式 $X_i \leqslant a, a < X_k < b (i<k<j)$ 及 $X_j \geqslant b$ 成立, 即 $V_{a,b}(N)$ 表示当 $N\leqslant n\leqslant 0$ 时 X_n 上穿区间 (a,b) 的次数, n 是从 N 到 0. 则由不等式 (4.11) 有

$$\begin{aligned} & E[V_{a,b}(N)] \\ \leqslant & (b-a)^{-1}\{E[(X_0-a)^+] - E[(X_N-a)^+]\} \\ \leqslant & (b-a)^{-1}E[(X_0-a)^+]. \end{aligned}$$

由于右端不取决于 N, 所以

$$E[V_{a,b}] \leqslant (b-a)^{-1}E[(X_0-a)^+],$$

其中 $V_{a,b} = V_{a,b}(-\infty)$ 是对于所有的 $n\leqslant 0, X_n$ 上穿 (a,b) 的次数.

作为这些加强不等式的推论, 指标集为 $\{\cdots,-2,-1,0\}$ 的鞅当 $n\to-\infty$ 时总有极限; 不需要附加假设,

$$X_{-\infty} = \lim_{n\to-\infty} X_n.$$

以概率 1 存在. 除此以外, 我们甚至可以得到更多的结果. 如果 $\{X_n; n=0,-1,-2,\cdots\}$ 是一个鞅, $\{\ |X_n|\ \}$ 是下鞅, 且由 (8.1)

314

$$\Pr\{W>\lambda\}\to 0, \quad \text{当}\lambda\to\infty,$$

其中 $W=\sup|X_n|$. 利用下鞅性质, 对任何 $\mathcal{F}_n$ 可测的事件 A_n 有 $E[\ |X_n|I(A_n)\ ] \leqslant E[\ |X_0|I(A_n)\ ]$, 因此我们得到

$$\begin{aligned} & \sup_{n\leqslant 0} E[\ |X_n|I\{\ |X_n|>c\ \}\ ] \\ \leqslant & \sup_{n\leqslant 0} E[\ |X_0|I\{\ |X_n|>c\ \}\ ] \\ \leqslant & E[\ |X_0|I\{W>c\}\ ]. \end{aligned}$$

同等式 (6.8) 一样的理由可证明最后一项当 $c \to \infty$ 时趋于 0. 于是, 鞅 $\{X_n; n = 0, -1, -2, \cdots\}$ 是一致可积的, 并且

$$\lim_{n \to -\infty} E[\,|X_n - X_{-\infty}|\,] = 0.$$

自然, 这个推理可以马上应用于以所有整数为指标集的鞅

$$\{X_n; n = \cdots, -1, 0, +1, \cdots\}.$$

对于这样的鞅.

$$X_{-\infty} = \lim_{n \to -\infty} X_n$$

始终存在, 并且有

$$E[\,|X_n - X_{-\infty}|\,] \to 0, \quad 当n \to -\infty时,$$

及

$$E[X_{-\infty}] = E[X_n] = E[X_0], 对于所有的n.$$

与此形成鲜明对比的是, 基本鞅收敛定理需要附加一些条件, 例如

$$\sup_{n \geqslant 0} E[X_n^+] < \infty,$$

本质上是为了保证极限

$$X_{+\infty} = \lim_{n \to +\infty} X_n$$

存在, 而要保证等式

$$E[X_{+\infty}] = E[X_0]$$

成立甚至需要更多的假设, 比如, 序列 $\{X_n; n \geqslant 0\}$ 是一致可积的.

设 $\{Z_n; n = 0, 1, \cdots\}$ 是概率空间 $(\Omega, \mathcal{F}, P)$ 上的随机变量序列, 并令 $\{\mathcal{G}_n; n = 0, 1, \cdots\}$ 是 $\mathcal{F}$ 的子 σ 域的递减序列, 即

$$\mathcal{F} \supset \mathcal{G}_n \supset \mathcal{G}_{n+1}, \qquad 对于所有的n.$$

315

称 $\{Z_n\}$ 为关于 $\{\mathcal{G}_n\}$ 的**反向鞅**, 如果对于 $n = 0, 1, \cdots$

(i) Z_n 是 $\mathcal{G}_n$可测的,

(ii) $E[\,|Z_n|\,] < \infty$, 且

(iii) $E[Z_n|\mathcal{G}_{n+1}] = Z_{n+1}$.

这样, $\{Z_n\}$ 是反向鞅, 当且仅当

$$X_n = Z_{-n}, \qquad n = 0, -1, -2, \cdots$$

构成一个关于 $\mathcal{F}_n = \mathcal{G}_{-n}, n = 0, -1, -2, \cdots$ 的鞅.

根据我们前面的讨论, 下面**反向鞅收敛定理**成立.

定理 8.1　设 $\{Z_n\}$ 关于 σ 域递减序列 $\{\mathcal{G}_n\}$ 是反向鞅. 则以概率 1 存在

$$Z = \lim_{n\to\infty} Z_n,$$

并且

$$\lim_{n\to\infty} E[\,|Z - Z_n|\,] = 0,$$

以及

$$E[Z_n] = E[Z], \qquad \text{对于所有的}n\text{成立}.$$

例　大数定律. 设 $X_1, X_2, \cdots$ 是独立同分布随机变量, 且 $E[\,|X_1|\,] < \infty$. 令 $\mu = E[X_1], S_0 = 0$, 当 $n \geqslant 1$ 时部分和 $S_n = X_1 + \cdots + X_n$. 令 $\mathcal{G}_n$ 是由 $\{S_n, S_{n+1}, \cdots\}$ 产生的 σ 域. 我们利用反向鞅理论导出强大数定律, 注意到

$$Z_n = n^{-1} S_n \qquad (Z_0 = \mu)$$

构成一个关于 $\mathcal{G}_n$ 的反向鞅. 很清楚, $E[\,|Z_n|\,] < \infty$, 并且 Z_n 是 $\mathcal{G}_n$可测的.

我们从显然的恒等式出发:

316

$$\begin{aligned} S_n &= E[S_n|S_n, S_{n+1}, \cdots] \\ &= \sum_{k=1}^{n} E[X_k|\mathcal{G}_n] \\ &= nE[X_k|\mathcal{G}_n] \qquad 1 \leqslant k \leqslant n, \end{aligned}$$

最后一个等号用到了被加项的对称性. 把上述关系式写成如下形式比较方便:

$$E[X_k|\mathcal{G}_n] = n^{-1} S_n = Z_n, \qquad 1 \leqslant k \leqslant n.$$

由此推得

$$\begin{aligned} E[Z_{n-1}|\mathcal{G}_n] &= (n-1)^{-1} E[S_{n-1}|\mathcal{G}_n] \\ &= (n-1)^{-1} \sum_{k=1}^{n-1} E[X_k|\mathcal{G}_n] \\ &= Z_n. \end{aligned}$$

这就验证了反向鞅的性质. [注: 为使 $\{Z_n\}$ 是反向鞅并不需要整个的独立性. 一个比较弱的充分条件是 $\{X_k\}$ 是**可交换**(也称为**对称**) 的随机变量**序列**, 这意味着 $(X_1, \cdots, X_n)$ 与 $(X_{\sigma(1)}, \cdots, X_{\sigma(n)})$ 对于每个自然数 n 具有相同的联合分布, 其中 $(\sigma(1), \cdots, \sigma(n))$ 是 $(1, \cdots, n)$ 的任意一个置换.]

应用反向鞅的收敛定理, 我们得到

$$Z = \lim_{n\to\infty} Z_n \qquad \text{以概率 1 存在,}$$

当且 $E[Z] = E[Z_n] = \mu$. 为了得到 Z 是非随机的, $X_1, X_2, \cdots$ 的独立性是极为重要的. 事实上, $Z \equiv \mu$. 证明如下. 对任何 $m = 1, 2, \cdots$,

$$Z = \lim_{n\to\infty} \frac{X_m + X_{m+1} + \cdots + X_{n+m}}{n},$$

因为有限项数的任意性并不影响极限. 由此推得, 对任何有限数 m, Z 和 $Z_m = m^{-1}S_m$ 是独立的, 因此对任意实数 a,

$$\Pr\{Z \geqslant a \text{ 且 } Z_m \geqslant a\} = \Pr\{Z \geqslant a\}\Pr\{Z_m \geqslant a\},$$

$$\Pr\{Z \geqslant a \text{且} \max_{n\leqslant k\leqslant m} Z_k \geqslant a\} = \Pr\{Z \geqslant a\}\Pr\{\max_{n\leqslant k\leqslant m} Z_k \geqslant a\},$$

并且

$$\begin{aligned}&\Pr\{Z \geqslant a \text{且} \limsup Z_n \geqslant a\}\\ &= \Pr\{Z \geqslant a\}\Pr\{\limsup Z_n \geqslant a\}.\end{aligned}$$

但 $Z = \lim Z_n = \limsup Z_n$, 所以

$$\Pr\{Z \geqslant a\} = [\Pr\{Z \geqslant a\}]^2.$$

317

由此推得对每个实数 a, $\Pr\{Z \geqslant a\}$ 只能取值 0 或 1, 并且这个性质蕴涵着 Z 是常数 (为什么?). 此外, 考虑到 $E[Z] = \mu$, 则 Z 的常数值一定是 μ. 这样我们就证明了强大数定理:

$$\lim_{n\to\infty} n^{-1}S_n = \mu$$

以概率 1 成立.

连续参数鞅

设 $\{X(t); t \geqslant 0\}$ 是概率空间 $(\Omega, \mathcal{F}, P)$ 上的连续参数随机过程. 对于每个 $t \geqslant 0$, 令 $\mathcal{F}_t$ 是 $\mathcal{F}$ 的子 σ 域, 并且

$$\mathcal{F}_s \subset \mathcal{F}_t, \qquad \text{如果} s \leqslant t.$$

在区间 $[0, \infty]$ 上取值的随机变量 T 称为关于 $\{\mathcal{F}_t\}$ 的马尔可夫时间, 如果对每个 $t \geqslant 0$, 事件 $\{T \leqslant t\}$ 属于 $\mathcal{F}_t$, 我们可以把 $\mathcal{F}_t$ 看作至时刻 t 为止得到的信息. 从这个观点讲, 马尔可夫时间小于或等于 t 的事件可由至时刻 t 为止得到的信息完全确定.

由于每个 σ 域包括它的任一集合的余集, 所以一个等价条件是

$$\{T > t\} \in \mathcal{F}_t, \qquad \text{对于所有的} t > 0. \tag{8.2}$$

对于连续参数过程, 只要求对每个 t, $\{T = t\}$ 是 $\mathcal{F}_t$ 中的事件是不够的. 然而, 同前面一样, 每个常数时间 $T \equiv \tau$ 是马尔可夫时间, 并且如果 S 和 T 是马尔可夫时间,

$$S + T, \quad S \wedge T = \min\{S, T\} \text{ 和 } S \vee T = \max\{S, T\}.$$

也都是马尔可夫时间. 这样, 如果 T 是马尔可夫时间, $T \wedge t = \min\{T, t\}$ 对每个固定 $t > 0$ 也是马尔可夫时间.

马尔可夫时间最重要的一个例子是过程首次到达水平 a 的时间 T_a, 即

$$T_a = \inf\{t \geqslant 0; X(t) \geqslant a\}.$$

现在我们假设每条轨道 $X(t)$ 是 t 的连续函数, 例如, 当 $X(t)$ 是布朗运动时, 就是这种情况. 令 $\mathcal{F}_t = \mathcal{F}(X(s); 0 \leqslant s \leqslant t)$ 是由 $\{X(s); 0 \leqslant s \leqslant t\}$ 生成的 σ 域. 则每个 $X(s)$ 是 $\mathcal{F}_s$ 可测的并且对于 $s < t, \mathcal{F}_s \subset \mathcal{F}_t$. 这时, T_a 是关于 $\{\mathcal{F}_t\}$ 的马尔可夫时间. 为了证明 (8.2), 注意到 $X(t)$ 是连续的, $\{T > t\}$ 这个事件的发生等价于对于某个 $k = 1, 2, \cdots$, 事件 $\left\{\min\limits_{0 \leqslant u \leqslant t}(a - X(u)) \geqslant 1/k\right\}$ 发生, 再一次利用连续性, 等价于对每个有理数 $r, 0 \leqslant r \leqslant t$, 事件 $\{(a - X(r)) \geqslant 1/k\}$ 同时发生. 即

318

$$\{T > t\} = \bigcup_{k=1}^{\infty} \bigcap_{\substack{0 \leqslant r \leqslant t \\ r \text{ 有理数}}} \{(a - X(r)) \geqslant 1/k\}.$$

当 $r \leqslant t$ 时, 每个事件 $\{(a - X(r)) \geqslant 1/k\} \in \mathcal{F}_t$. 因为 $\mathcal{F}_t$ 是 σ 域, 并且 $[0, t]$ 中的有理数 r 是可数的, 所以

$$\bigcap_{0 \leqslant r \leqslant t} \{(a - X(r)) \geqslant 1/k\} \in \mathcal{F}_t.$$

再者, $\mathcal{F}_t$ 中可数多集合的并集仍然属于 $\mathcal{F}_t$, 因此 $\{T > t\} \in \mathcal{F}_t$, 这正是要证明的.

设 A 是闭集, 定义 A 的**初遇时 $T(A)$**为

$$T(A) = \inf\{t \geqslant 0; X(t) \in A\}.$$

同时可证 $T(A)$ 是关于 $\mathcal{F}_t = \mathcal{F}(X(u); 0 \leqslant u \leqslant t)$ 的马尔可夫时间, 只要 $X(t)$ 是连续函数.

不幸的是, 由于一些技术上的原因, 如果 $X(t)$ 不是连续的或 A 不是闭的, 集合 A 的初遇时不一定是关于 $\{\mathcal{F}_t\}$ 的马尔可夫时间. 然而, 我们有可能通过扩大 σ 域

$\mathcal{F}_t$ 来克服这个缺点. 假设每一个实现 $X(t)$ 作为 t 的函数是从右边连续的并且具有左极限. 即, 假设

$$X(t) = \lim_{s \downarrow t} X(s), \qquad \text{对于所有的 } t \geqslant 0,$$

和

$$X(t-) = \lim_{s \uparrow t} X(s), \qquad \text{对于所有的 } t > 0 \text{存在}.$$

依然是 $\mathcal{F}_t = \mathcal{F}(X(u); 0 \leqslant u \leqslant t)$, 令 $\mathcal{F}_{t+}$ 是由同时属于每一个 σ 域 $\mathcal{F}_{t+\varepsilon}$ 的事件组成的, 其中 $\varepsilon > 0$ 是任意的. 用集合论语言来表示, $\mathcal{F}_{t+}$ 就是交集

$$\mathcal{F}_{t+} = \bigcap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon}.$$

$\mathcal{F}_{t+}$ 是一个 σ 域, $X(t)$ 是 $\mathcal{F}_{t+}$ 可测的, 并且

$$\mathcal{F}_{s+} \subset \mathcal{F}_{t+}, \qquad \text{如果 } s < t.$$

最后, 令 $\bar{\mathcal{F}}_{t+}$ 表示包含 $\mathcal{F}_{t+}$ 的每个集合和 Ω 中 $A \in \mathcal{F}$ 的一切子集的最小 σ 域, 这里 $P[A] = 0$. 粗略地说, $\bar{\mathcal{F}}_{t+}$ 是由所有那些与 $\mathcal{F}_{t+}$ 中事件概率等价的事件组成.

这样, 对于每个 Borel 集 A, 初遇时

$$T(A) = \begin{cases} \inf\{t \geqslant 0 : X(t) \in A\}, & \text{如果 } X(t) \in A \text{ 对某个 } t \geqslant 0, \\ \infty, & \text{如果 } X(t) \notin A \text{ 对于所有的 } t, \end{cases}$$

是关于 $\{\bar{\mathcal{F}}_{t+}\}$ 的马尔可夫时间.

319

关于鞅的可选抽样定理和收敛定理在连续时间情况仍是正确的. 如果 $\{X(t); t \geqslant 0\}$ 是关于 $\{\mathcal{F}_t\}$ 的下鞅, 则对于所有的马尔可夫时间 T

$$E[X(0)] \leqslant E[X(T \wedge t)] \leqslant E[X(t)], \quad t \geqslant 0. \tag{8.3}$$

对于上鞅的情况, 上述不等号是相反的, 对于鞅的情况, 上述等号成立. 如果 $\Pr\{T < \infty\} = 1$, 则

$$X(T \wedge t) \to X(T), \qquad \text{当 } t \to \infty.$$

如果我们能验证此极限与期望在 (8.3) 中可交换, 便可推得可选抽样定理.

定理 8.2 设 $\{X(t); t \geqslant 0\}$ 是下鞅, T 是关于 $\{\mathcal{F}_t\}$ 的马尔可夫时间. 如果 $\Pr\{T < \infty\} = 1$, 并且随机变量 $\{X(t \wedge T)^+; t \geqslant 0\}$ 是一致可积的, 则

$$E[X(0)] \leqslant E[X(T)].$$

推论 8.1 设 $\{X(t); t \geqslant 0\}$ 是鞅, 而 T 是马尔可夫时间, 如果 $\Pr\{T < \infty\} = 1$ 并且 $E[\sup_{t \geqslant 0} |X(t)|\,] < \infty$, 则

$$E[X(0)] = E[X(T)].$$

在第 7 章我们将利用这些结果导出布朗运动的若干重要性质. 如果 $\{X(t);t\geqslant 0\}$ 是布朗运动过程, 其均值为 0, 方差为 σ^2, 则下面三个过程

(i) $X(t)$,

(ii) $Y(t)=X^2(t)-\sigma^2 t$,

(iii) $Z(t)=\exp\{\theta X(t)-\frac{1}{2}\theta^2\sigma^2 t\}$, θ 为实数,

都是关于 $\mathcal{F}_t=\mathcal{F}(X(u);0\leqslant u\leqslant t)$ 的鞅. 这将在 7.5 节证明. 在那里, 这些鞅将被用来计算与布朗运动相联系的若干重要的量.

可以证明 (但证明相当困难) 对于实数 θ 和具有一阶、二阶连续导数的严格增函数 f, 假如我们规定 $E[W(t)]<\infty$, 则

$$W(t)=\exp\left\{\theta f[X(t)]-\frac{\sigma^2}{2}\int_0^t\{\theta^2(f'[X(u)])^2+\theta f''[X(u)]\}\mathrm{d}u\right\}$$

320 是一个鞅.

泊松过程

如果 $\{X(t);t\geqslant 0\}$ 是具有参数 λ 的泊松过程, 则

$$Y(t)=X(t)-\lambda t, \tag{8.4}$$

$$U(t)=Y^2(t)-\lambda t, \tag{8.5}$$

和

$$V(t)=\exp\{-\theta X(t)+\lambda t(1-\mathrm{e}^{-\theta})\},\quad -\infty<\theta<\infty, \tag{8.6}$$

是关于 $\mathcal{F}_t=\mathcal{F}(X(u);0\leqslant u\leqslant t)$ 的鞅.

固定一个正整数 a, 并令 T_a 是 $X(t)$ 首次达到 a 的时间. 假设 $X(0)=0$ 并注意到 $X(T_a)=a$(因为泊松过程的变化是每次跳跃一个单位), 应用可选抽样定理于 (8.4)~(8.6), 我们得到 $a=\lambda E[T_a]$,

$$E[(a-\lambda T_a)^2]=\lambda E[T_a]=a\ \text{或方差}\ \mathrm{Var}[T_a]=a/\lambda^2,$$

以及

$$\mathrm{e}^{\theta a}=E[\exp\{-\beta T_a\}],$$

或

$$E[\exp\{-\beta T_a\}]=\left(\frac{\lambda}{\lambda+\beta}\right)^a.$$

其中 $\beta=-\lambda(1-\mathrm{e}^{-\theta})$. 最后的表达式是 T_a 分布的拉普拉斯变换. 正如我们已经知道的, 它证明了 T_a 具有参数为 a 和 λ 的 Γ 分布.

纯生过程

若 $\{X(t);t\geqslant 0\}$ 是具有增长参数 $\lambda(i)\geqslant 0\quad(i\geqslant 0)$ 的纯生过程. 为方便起见设 $X(0)=0$. 我们断言

$$Y(t)=X(t)-\int_0^t\lambda[X(u)]\mathrm{d}u,$$

和

$$V(t)=\exp\{\theta X(t)+[1-\mathrm{e}^{\theta}]\int_0^t\lambda[X(u)]\mathrm{d}u\}\tag{8.7}$$

都是鞅, 其中 θ 是固定的, 假如它们的期望都是有限的. 有许多方法可以验证这些结论. 最好的方法是把该问题化为关于泊松过程的相应问题. 321

令 $\tau_0,\tau_1,\cdots$ 表示在给定过程中相继两个增殖的间隔时间. τ_i 是独立的, 并且 τ_k 具有参数为 $\lambda(k)$ 的指数分布. 随机变量 $\sigma_k=\lambda(k)\tau_k,k=0,1,\cdots$ 仍然是独立的. 此外还具有参数为 1 的共同指数分布. 它们可以作为标准泊松过程 $\{N(t);t\geqslant 0\}$ 中的间隔时间. $X(t)$ 和 $N(t)$ 之间的关系可以用图 6-1 解释.

由前所述

$$W(T)=\exp\{\theta N(T)+T(1-\mathrm{e}^{\theta})\},\qquad T\geqslant 0$$

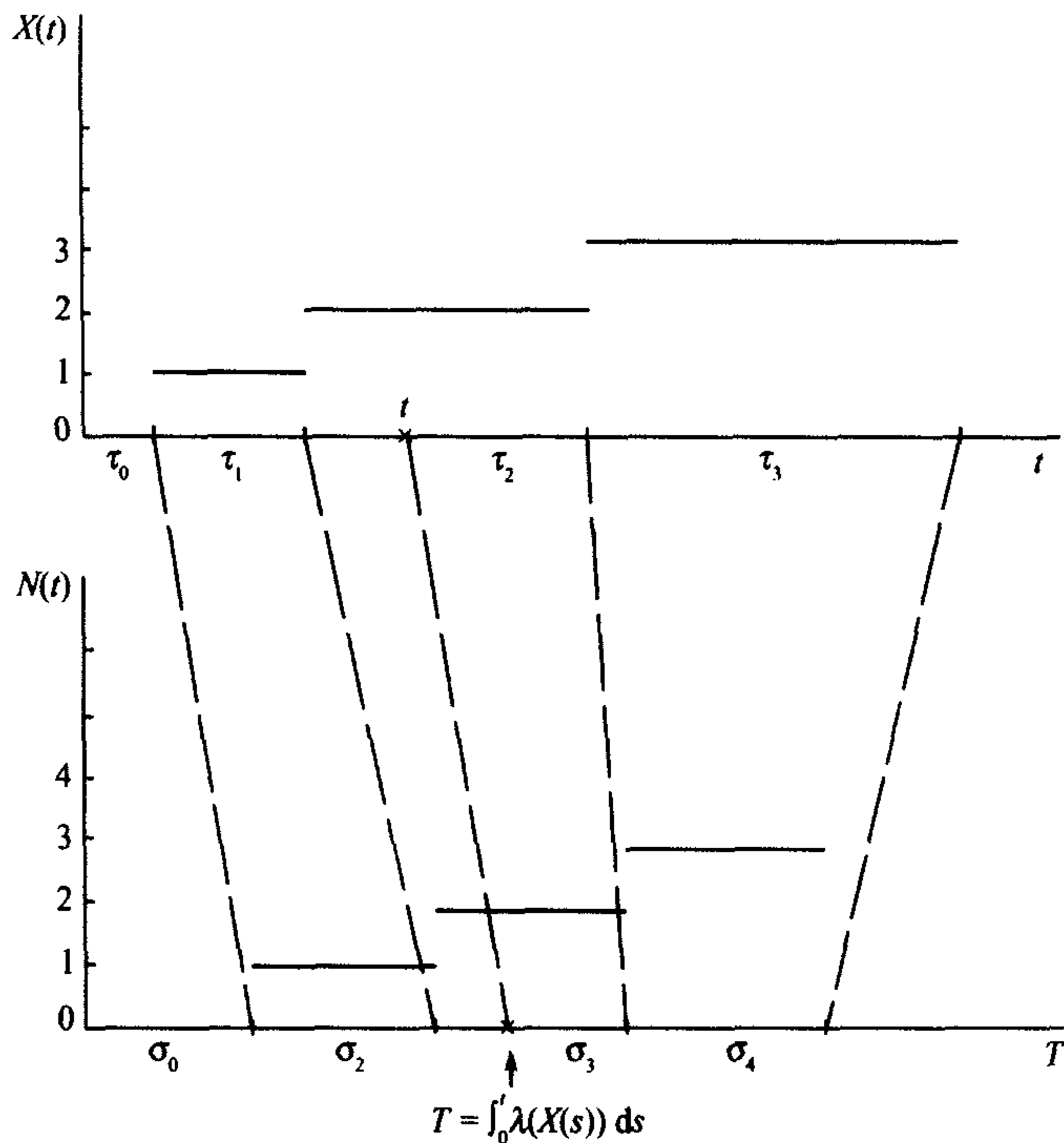


图 6-1 所有间隔时间 $\sigma_k=\lambda(k)\tau_k$ 的均值都为 1, 因此定义了一个泊松过程 322

是关于 $\mathcal{G}_T = \mathcal{F}(N(u); 0 \leqslant u \leqslant T)$ 的鞅. 于是, 对于 $T \geqslant S$,

$$E[W(T)|N(u); 0 \leqslant u \leqslant S] = W(S).$$

固定点 t 并令 $T = T(t)$ 是 T 坐标轴上的对应点. 其关系是

$$T = T(t) = \int_0^t \lambda[X(u)]\mathrm{d}u, \tag{8.8}$$

且

$$N(T) = X(t). \tag{8.9}$$

固定点 $s < t$, 并令 $S = S(s)$ 以类似方式对应于它. 关于 $\{N(u); 0 \leqslant u \leqslant S\}$ 的条件与关于 $\{X(u); 0 \leqslant u \leqslant s\}$ 的条件是等价的. 这样,

$$E[W(T)|X(u); 0 \leqslant u \leqslant s] = W(S). \tag{8.10}$$

但是通过替换 (8.8) 和 (8.9), 我们记 $W(T) = V(t)$ 和 $W(S) = V(s)$. 这样,

$$E[V(t)|\mathcal{F}_s] = V(s)$$

其中 $\mathcal{F}_s = \mathcal{F}(X(u); 0 \leqslant u \leqslant s)$, 因此 $\{V(t)\}$ 是一个鞅.

细心的读者会注意到我们讨论中的一个漏洞. 对固定的 $t, T = T(t)$ 不是固定的, 而是随机的. 但是 T 是关于 $\{\mathcal{G}_T\}$ 的马尔可夫时间, 所以应用可选抽样定理的推广即可验证 (8.10).

形式上, 我们可以证明 $Y(t) = X(t) - \int_0^t \lambda[X(u)]\mathrm{d}u$ 是一个鞅, 这只要对下式

$$\theta^{-1}[V(t) - 1] = Y(t) + o(\theta)$$

令 θ 趋于 0 即得到, 其中 $o(\theta)$ 是 (随机) 余项, 左边部分对每个 $\theta \neq 0$ 都是鞅, 因此 $Y(t)$ 也是一个鞅.

生灭过程

设 $\{X(t); t \geqslant 0\}$ 是生灭过程, 其增长参数为 $\lambda_i = \lambda(i),\quad i \geqslant 0$, 消亡参数为 $\mu_i = \mu(i), i \geqslant 1$. 假设 $\lambda(0) = 0$, 所以 0 是吸收状态, 但对于 $i \geqslant 1$, 假设 $\lambda(i) > 0$. 定义

$$f(0) = 0, \qquad f(1) = 1,$$

及

323

$$f(j) = 1 + \frac{\mu_1}{\lambda_1} + \frac{\mu_1\mu_2}{\lambda_1\lambda_2} + \cdots + \frac{\mu_1 \cdots \mu_{j-1}}{\lambda_1 \cdots \lambda_{j-1}}, \text{对} j \geqslant 2. \tag{8.11}$$

那么 $Z(t)=f[X(t)]$ 是一个鞅, 只要它的均值是有限的. (试与初等问题 25 比较.) 为此, 固定 $s<t$ 和某状态 $i\geqslant 1$, 并考虑

$$\begin{aligned}g_i(t)&=E[Z(t)|X(u);0\leqslant u\leqslant s,X(s)=i]\\&=E[Z(t)|X(s)=i],\end{aligned}$$

最后这个等式由马尔可夫性推得. 那么, 对于较小的 $h>0$, 通过考虑在时间区间 $(t,t+h)$ 内可能出现的转移, 我们得到等式

$$\begin{aligned}g_i(t+h)&=\sum_{k=0}^{\infty}E[Z(t+h)|X(t)=k]\Pr\{X(t)=k|X(s)=i\}\\&=g_i(t)+h\sum_{k=0}^{\infty}\{\lambda_k[f(k+1)-f(k)]-\mu_k[f(k)-f(k-1)]\}\cdot\\&\quad\Pr\{X(t)=k|X(s)=i\}+o(h).\end{aligned}$$

这样

$$\begin{aligned}g_i'(t)&=\lim_{h\downarrow 0}\frac{g_i(t+h)-g_i(t)}{h}\\&=\sum_{k=0}^{\infty}\{\lambda_k[f(k+1)-f(k)]-\mu_k[f(k)-f(k-1)]\}\Pr\{X(t)=k|X(s)=i\}\\&=\sum_{k=0}^{\infty}\left\{\lambda_k\left[\frac{\mu_1\cdots\mu_k}{\lambda_1\cdots\lambda_k}\right]-\mu_k\left[\frac{\mu_1\cdots\mu_{k-1}}{\lambda_1\cdots\lambda_{k-1}}\right]\right\}\Pr\{X(t)=k|X(s)=i\}\\&=0\end{aligned}$$

由于 $g_i'(t)=0, g_i(t)=E[Z(t)|X(s)=i]$ 对于 $t>s$ 是 t 的常数函数, 令 $t\downarrow s$, 我们得到

$$\begin{aligned}g_i(s)&=E[Z(s)|X(s)=i]\\&=g_i(t)=E[Z(t)|X(s)=i],\end{aligned}$$

即 $Z(t)$ 是一个鞅.

固定状态 $i<m$ 并令 $v(i)$ 是在条件 $X(0)=i$ 下过程到达状态 m 之前被 0 状态吸收的概率. 令

$$T_{0,m}=\inf\{t\geqslant 0: X(t)=0 \text{ 或 } X(t)=m\}.$$

应用可选抽样定理, 得

$$f(i)=E[Z(T_{0,m})]=v(i)\cdot 0+(1-v(i))f(m),$$

因此

$$v(i)=\frac{f(m)-f(i)}{f(m)},$$

其中 f 如 (8.11) 定义.

这里, 由不同的方法得到了一个类似于第 4 章定理 7.1 的结果.

324

有许多与生灭过程关联的鞅, 我们叙述其中的两个:

(a) 设 $g(i), i=0,1,\cdots$ 使得

$$Y(t)=g[X(t)]-\int_0^t\{\lambda[X(u)][g(X(u)+1)-g(X(u))]-\mu[X(u)][g(X(u))-g(X(u)-1)]\}\mathrm{d}u$$

的期望是有限的, 则 $\{Y(t)\}$ 是一个鞅.

注意, 当 $g(i)$ 是

$$\lambda(i)[g(i+1)-g(i)]-\mu(i)[g(i)-g(i-1)]\equiv 1,\quad i\geqslant 1$$

的解时, 上述积分可以大大简化.

(b) 设 $g(i), i=0,1,\cdots$ 使得对于某个固定的实数 θ,

$$\begin{aligned}V(t)=\exp(&-\theta g[X(t)]\\&-\int_0^t[\lambda(X(u))\{1-\mathrm{e}^{-\theta[g(X(u)+1)-g(X(u))]}\}+\\&\mu(X(u))\{1-\mathrm{e}^{+\theta[g(X(u))-g(X(u)-1)]}\}]\mathrm{d}u)\end{aligned}$$

的期望是有限的, 则 $\{V(t)\}$ 是鞅.

初等问题

1. 考虑在平面第一象限整数格点上的随机游动. 如果过程在某一步停在 (m,n), 则下一步将以概率 $\frac{1}{2}$ 转移至 $(m+1,n)$ 或 $(m,n+1)$. 设过程从 $(0,0)$ 出发, Γ 是第一象限内联结相邻格点 (从 Y 轴到达 X 轴) 的任意一条曲线. Y_1 和 Y_2 分别表示过程在到达边界 Γ 之前向右和向上的步数. 试证明 $EY_1=EY_2$. 图 6-2 描出了曲线 Γ 的一个例子.

提示: 利用对部分和的可选停止定理证明 $E[Y_1]=E[Y_2]=\frac{1}{2}E[T]$, 其中 T 是过程到达边界 Γ 所需要的步数.

325

2. 考虑下面离散时间马尔可夫过程, 其状态空间为单位区间. 如果过程现在位于 $p(0<p<1)$, 下一次试验它将以概率 p 跳到 $\alpha+\beta p$, 以概率 $1-p$ 跳到 βp, 其中 $\alpha,\beta>0$, 且 $\alpha+\beta=1$. 用公式来表达此过程是

$$X_{n+1}=\begin{cases}\alpha+\beta X_n, & \text{以概率 } X_n,\\ \beta X_n, & \text{以概率 } 1-X_n.\end{cases}$$

试证明这个过程是一个鞅.

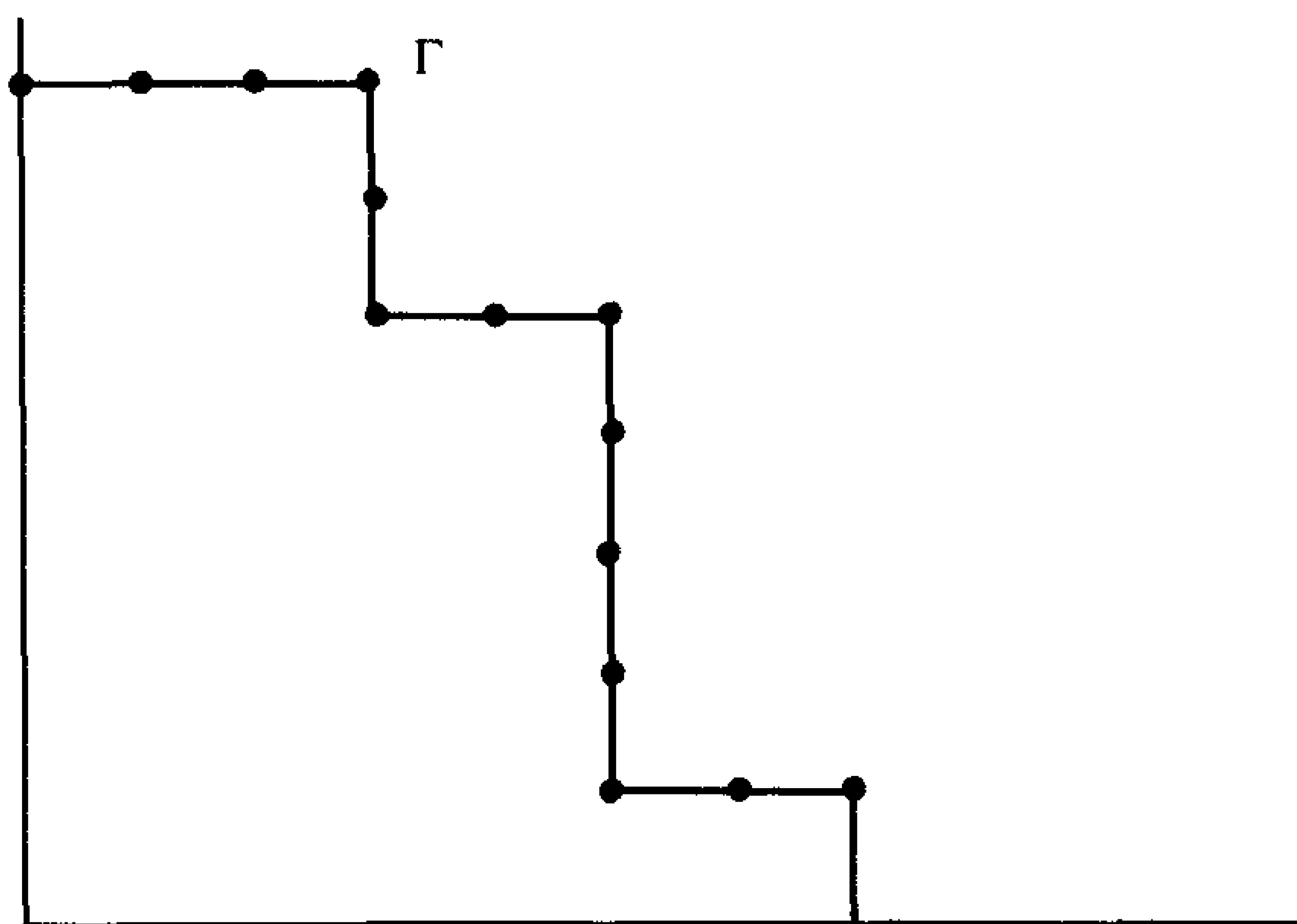


图 6-2

3. 设 $W(n)$ 是带有迁入的分支过程:

$$W(n+1) = Y_n + X_{n,1} + X_{n,2} + \cdots + X_{n,W(n)},$$

其中 Y_n 表示第 n 代的迁入, $X_{n,j}$ 表示第 n 代的第 j 个体后代数目, 它们都相互独立. 假设 $E[Y_n] = \lambda$ 和 $E[X_{n,j}] = m \neq 1$, 证明

$$Z_n = m^{-n}\left[W(n) - \lambda\left\{\frac{1-m^n}{1-m}\right\}\right]$$

是一个鞅.

4. 设某群体第 n 代男性数目为 X_n, 第 n 代女性数目是 Y_n. 假定男女配对是稳定的, 这样有 $Z_n = \min\{X_n, Y_n\}$ 对产生后代, 各对相互独立产生后代并都有相同母函数 $g(t,s) = E[t^{\xi}s^{\eta}]$, 其中 ξ 表示单独一对双亲所具有的男性后代的数目, η 表示单独一对双亲所具有的女性后代的数目. 证明, 若 $E[\xi] \leqslant 1$ 或 $E[\eta] \leqslant 1$, 则 $\{Z_n\}$ 是一个上鞅.

5. 假设 $Y_1, Y_2, \cdots$ 是独立同分布随机变量序列, $\Pr\{Y_1 = +1\} = p$, $\Pr\{Y_1 = -1\} = q = 1-p$. 固定正整数 a 和 b, 记 $S_0 = 0, S_n = Y_1 + \cdots + Y_n, n \geqslant 1$, 令

$$T = \min\{n : S_n = -a \text{ 或 } S_n = b\}.$$

试证明

$$E[T] = \frac{b}{p-q} - \frac{a+b}{p-q} \cdot \frac{1-(p/q)^b}{1-(p/q)^{a+b}}, \quad \text{当} p \neq q.$$

(当 $p = q = \dfrac{1}{2}$ 时, $E[T]=ab$ 已在 6.4 节例 (a) 中导出.)

6. 设 $Y_1, Y_2, \cdots$ 是独立同分布且 $\Pr\{Y_1=+1\}=p, \Pr\{Y_1=-1\}=q=1-p$. 假设 $p>\frac{1}{2}>q$：置 $S_0=0$ 且 $S_n=Y_1+\cdots+Y_n(n\geqslant 1)$. 对某个固定正整数 b, 令

326

$$T=\min\{n: S_n\geqslant b\}.$$

试证其母函数为

$$E[s^T]=\left(\frac{1-\{1-4pqs^2\}^{1/2}}{2qs}\right)^b, \quad 0<s\leqslant 1.$$

提示：利用 Wald 等式.

7. 在初等问题 6 的条件下, 试证明均值 $E[T]=b/(p-q)$, 方差 $\mathrm{Var}[T]=b[1-(p-q)^2]/(p-q)^3$.

8. 设 $Y(0), Y(1), \cdots$ 是成功游程马尔可夫链, 其中 $P_{00}=1$, 因此 0 是一个吸收壁, 且 $P_{i,i+1}=p, P_{i0}=q=1-p, i=1,2,\cdots$. 设 a 和 b 是任意常数, 试证明

$$X_n=\begin{cases} b, & \text{若 } Y(n)=0, \\ a\,(1/p)^{Y(n)-1}+b[1-(1/p)^{Y(n)-1}], & \text{若 } Y(n)>0, \end{cases}$$

是一个鞅.

9. 考虑随机变量族 $\{X_n\}_0^\infty$, 其中每个随机变量绝对值的期望是有限的, 并且满足关系式

$$E[X_{n+1}|X_0, X_1, \cdots, X_n]=\alpha X_n+\beta X_{n-1}, n>0,$$

其中 $\alpha>0, \beta>0, \alpha+\beta=1$. 试求一适当值 a 使得序列

$$Y_n=aX_n+X_{n-1}, n\geqslant 1, Y_0=X_0,$$

关于 $\{X_n\}$ 构成一个鞅.

10. 设 $\{X_n; n\geqslant 0\}$ 关于 $\{Y_n\}$ 是一个鞅. 证明对于任何满足关系式 $k\leqslant l<m$ 的正整数集合, 若 X_m-X_l 与 X_k 是无关的, 即

$$E[(X_m-X_l)X_k]=0.$$

提示：关于条件 $Y_1, Y_2, \cdots, Y_k$ 求期望.

11. 设 $\{\xi_i\}$ 是这样一个随机变量序列, 它的部分和

$$X_n=\xi_0+\xi_1+\cdots+\xi_n, \qquad n\geqslant 1,$$

构成一个鞅. 试证明被加项必定互不相关, 即 $E[\xi_i\xi_j]=0, \quad i\neq j$.

12. 设 $S_n = X_1 + \cdots + X_n$ 是一个鞅, 且对于所有的 k 满足 $E[X_k^2] \leqslant K < \infty$. 试证明 S_n 服从弱大数律: 对任给正数 ξ

$$\Pr\{|S_n/n| > \varepsilon\} \to 0, \quad 当n \to \infty.$$

提示: 利用最大值不等式和初等问题 11 的正交性结论.

327

13. 设 $\{\xi_i\}_{i=0}^{\infty}$ 是一实值随机变量序列, 满足 $E[\xi_i|\xi_0, \xi_1, \cdots, \xi_{i-1}] = 0, i = 1, 2, \cdots$. 定义

$$X_0 = \xi_0, X_{n+1} = \sum_{i=0}^{n} \xi_{i+1} f_i(\xi_0, \xi_1, \cdots, \xi_i),$$

其中 f_i 是给定的 $i+1$ 元实函数序列. 试证明 $\{X_n\}$ 构成一个鞅.

14. 考虑重复独立地抛掷一枚均匀硬币的赌博, 其中任意 k 轮的结果 ξ_k 满足 $\Pr\{\xi_k = 1\} = \Pr\{\xi_k = -1\} = \frac{1}{2}$. 假如一个赌徒在第一输中下赌注一个单位, 此后若输了就下赌注所输的一倍, 若赢了就下赌注一个单位. 假设赌徒有无限财富. 令 X_n 表示在第 n 输之后净赢得数值. 试证明 $\{X_n\}_1^{\infty}$ 关于 $\{\xi_n\}_1^{\infty}$ 是一个鞅.

提示: 对适当的 f_k, X_n 可表示为如下形式

$$X_n = \sum_{k=1}^{n} \xi_k f_k(\xi_1, \cdots, \xi_{k-1}),$$

并参考初等问题 13.

15. (a) 考虑状态空间为 $\{0, 1, 2, \cdots, N\}$ 上的马尔可夫链 $\{X_n; n > 0\}$, 其转移概率矩阵为

$$(*) \qquad P_{ij} = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}.$$

试证明 $\{X_n; n \geqslant 0\}$ 和 $\left\{V_n = \dfrac{X_n(N - X_n)}{(1 - N^{-1})^n}, \quad n \geqslant 0\right\}$ 都关于 $\{X_n\}$ 构成鞅.

(b) 用

$$(**) \qquad P_{ij} = \frac{\dbinom{2i}{j}\dbinom{2N-2i}{N-j}}{\dbinom{2i}{N}}$$

代替 $(*)$, 试确定 λ, 使得

$$W_n = \frac{X_n(N - X_n)}{\lambda^n}, \quad n \geqslant 0$$

关于 $\{X_n\}$ 是一个鞅.

16. 假设 Y_0 在 $(0,1]$ 上均匀分布, 给定 Y_n, 假定 Y_{n+1} 也在 $(1-Y_n,1]$ 上均匀分布. 试证 $X_0=Y_0$ 及

$$X_n = 2^n \prod_{k=1}^{n}\left[\frac{1-Y_k}{Y_{k-1}}\right], \qquad n=1,2,\cdots,$$

328 是一个鞅.

17. 一个罐在初始时刻装有一个红球和一个绿球. 随机地抽取一个球, 然后把它及另外一个同样颜色的球再放入罐中. 这个步骤重复无限多次. 令 X_n 表示在第 n 步红球所占的比例. 则 $\{X_n\}$ 是一个鞅 [参见 6.1 节例 (i)]. (a) 试利用最大值不等式证明 $\Pr\{X_n \geqslant 3/4$对于某个$n=1,2,\cdots\} \leqslant 2/3$. 换句话说, 红球所占比例超过 3/4 的概率无论为何不会超过 2/3. (b) 利用 6.6 节例 (c) 中的极限分布证明 $\lim_{n\to\infty}\Pr\{X_n \geqslant 3/4\}=1/4$. 即在极限情况下, 红球比例不小于 3/4 的概率是 1/4.

18. 设 $P_{ij}=\mathrm{e}^{-i}i^j/j!, i,j=0,1,\cdots$ 是马尔可夫链 X_n 的转移概率, 其中规定 $P_{00}=1$. (a) 试验证 X_n 是一个鞅. (b) 试导出不等式

$$\Pr\left\{\max_{0\leqslant n<\infty} X_n \geqslant a | X_0=i\right\} \leqslant i/a,$$

其中 $i,a=1,2,\cdots$.(c) 试证明以概率 1 $\lim_{n\to\infty} X_n=0$.

提示: 利用可选停止定理, 其中 T 是使得 $X_n=0$ 或 $X_n \geqslant a$ 的首达时 n.

19. 设 X_n 是马尔可夫链, 其转移概率是 $P_{ij}=1/[\mathrm{e}(j-i)!](i=0,1,\cdots,j=i,i+1,\cdots)$. 证明下面三个过程都是鞅:

(a) $Y_n=X_n-n$;

(b) $U_n=Y_n^2-n$;

(c) $V_n=\exp\{X_n-n(\mathrm{e}-1)\}$.

20. 设 $\{X_n\}_1^\infty$ 是下鞅. 证明序列

$$U_1=0, U_n=\sum_{i=2}^{n}\{E[X_i|X_1,\cdots,X_{i-1}]-X_{i-1}\}, n\geqslant 2,$$

是一个递增过程, 即 $U_n \geqslant U_{n-1}$.

21. 考虑马尔可夫链 $\{X_n; n\geqslant 0\}$, 其状态空间为非负整数, 转移概率矩阵为 $\boldsymbol{P}=\| P_{ij} \|$. 设 $u(i,n)$ 是定义在整数 $i,n\geqslant 0$ 上一个函数且满足函数方程

$$u(i,n)=\sum_{k=0}^{\infty} u(k,n+m)P_{ik}^{(m)},$$

其中 $P_{ik}^{(m)}$ 是从状态 i 到状态 k 的 m 步转移概率. 试证明

$$U_n=n(X_n,n)$$

关于 $\{X_n\}$ 是一个鞅.

22. 考虑一具有 $N+1$ 个状态的马尔可夫链 $\{X_n; n \geqslant 0\}$, 其可能状态值是 $x_0 < x_1 < \cdots < x_N$, 转移概率矩阵为 $P_{ij} = \Pr\{X_{n+1} = x_j | X_n = x_i\}$. 假设 $\{X_n\}$ 也是一个鞅, 试证明状态 x_0 和 x_N 是吸收的, 即 $P_{0,0} = P_{N,N} = 1$.

329

23. 考虑一个具有 $N+1$ 个状态 $\{0, 1, \cdots, N\}$ 的马尔可夫链 $\{X_n; n \geqslant 0\}$, 其转移概率

$$P_{ij} = \begin{pmatrix} N \\ j \end{pmatrix} \pi_i^j (1-\pi_i)^{N-j},$$

其中

$$\pi_i = \frac{1 - \mathrm{e}^{-2ai/N}}{1 - \mathrm{e}^{-2a}}.$$

试证明 $Z_n = \mathrm{e}^{-2aX_n}$ 是一个鞅.

24. 设 $\{X_n, n \geqslant 0\}$ 是状态空间为非负整数, 转移概率矩阵为 $\boldsymbol{P} = \| P_{ij} \|$ 的瞬态马尔可夫链. 定义

$$u(i) = \sum_{n=0}^{\infty} P_{i0}^{(n)}.$$

试证明 $U_n = u(X_n)$ 是下鞅.

25. 考虑一个具有无穷小参数 λ_i 和 μ_i 的有限生灭过程 $\{X(t), t \geqslant 0\}, 0 \leqslant i \leqslant N (\mu_0 = 0)$. 无穷小矩阵是

$$\boldsymbol{A} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & 0 \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ . & . & . & \mu_N & -\mu_N \end{pmatrix}.$$

考虑线性方程组

$$\boldsymbol{A y} = \boldsymbol{0}$$

的任意一个解 $\boldsymbol{y} = (y_0, y_1, \cdots)$. 试证明 $Y(t) = y_{X(t)}, t > 0$, 关于 $\mathcal{F}_t = \sigma(X(u); u \leqslant t)$ 是一个鞅.

提示: 证明若 $\boldsymbol{A y} = \boldsymbol{0}$, 则

$$y_i = \sum_{j=0}^{N} P_{ij}(t) y_j, \qquad i = 0, 1, \cdots, N,$$

对于所有的 $t > 0$ 成立, 其中 $P_{ij}(t)$ 表示过程 $\{X(t); t \geqslant 0\}$ 转移矩阵.

问 题

1. 试证明：若 $\{X_n\}$ 是下鞅. $\varphi(x)$ 是凸的增函数, 且对于任意的 n 有 $E|\varphi^+(X_n)| < \infty$, 则 $\{\varphi(X_n)\}$ 是下鞅. (参考引理 2.2.)

330

2. 设 $\boldsymbol{P} = \| P_{ij} \|$ 是不可约常返马尔可夫链 $\{X_n\}$ 的转移概率矩阵. 试利用下鞅收敛定理 (见注 5.1) 证明对于所有的 i 不等式组

$$y(i) \geqslant \sum_{j=0}^{\infty} P_{ij} y(j)$$

的每个非负解 $y = \{y(i)\}$ 是一个常数.

提示：与 6.6 节例 (a) 方法相同.

3. 设 $\{U_n\}$ 和 $\{V_n\}$ 关于相同的过程 $\{Y_n\}$ 为鞅, 并且 $U_0 = V_0 = 0, E[U_n^2] < \infty, E[V_n^2] < \infty$, 对于所有的 n 成立. 试证明

$$E[U_n V_n] = \sum_{k=1}^{n} E[(U_k - U_{k-1})(V_k - V_{k-1})].$$

特别地, 有

$$E[U_n^2] = \sum_{k=1}^{n} E[(U_k - U_{k-1})^2].$$

提示：由于 $U_n V_n = \sum_{k=1}^{n} (U_k V_k - U_{k-1} V_{k-1})$, 故只要证明 $E[U_k V_k - U_{k-1} V_{k-1}] = E[(U_k - U_{k-1})(V_k - V_{k-1})]$. 但是 $E[(U_k - U_{k-1})(V_k - V_{k-1})] = E[U_k V_k] - E[U_{k-1} V_k] - E[(U_k - U_{k-1}) V_{k-1}]$. 然后, 由对 $Y_0, \cdots, Y_{k-1}$ 的限制条件并利用鞅的性质, 求最后两个期望.

4. 设 $\{X_n\}$ 是鞅, 且对于某个 $\alpha > 1$ 满足

$$E[\, |X_n|^\alpha \,] < \infty, \qquad \text{对于所有的} n.$$

试证明

$$E\left[\max_{0 \leqslant k \leqslant n} |X_k|\right] \leqslant \frac{\alpha}{\alpha - 1} E[\, |X_n|^\alpha \,]^{1/\alpha}.$$

提示：$E\left[\max_{0 \leqslant k \leqslant n} |X_k|\right] = \int_0^\infty \Pr\left\{\max_{0 \leqslant k \leqslant n} |X_k| > t\right\} \mathrm{d}t$. 然后对下鞅 $\{\, |X_n|^\alpha \,\}$ 利用最大值不等式.

5. 设 $\{X_n\}$ 是下鞅. 试证明加强最大值不等式

$$\lambda \Pr\left\{\max_{0\leqslant k\leqslant n} X_k > \lambda\right\} \leqslant E\left[X_n I\left\{\max_{0\leqslant k\leqslant n} X_k > \lambda\right\}\right]$$
$$\leqslant E[X_n^+] \leqslant E[\,|X_n|\,], \qquad \lambda > 0.$$

(注意: 引理 5.1 要求对于所有的 k 有 $X_k \geqslant 0$, 但这里并无此要求.) 因此, 对于一个鞅 $\{X_n\}$

$$\lambda \Pr\left\{\max_{0\leqslant k\leqslant n} |X_k| > \lambda\right\} \leqslant E\left[\,|X_n| I\left\{\max_{0\leqslant k\leqslant n} |X_k| > \lambda\right\}\right], \quad \lambda > 0.$$

331

6. 问题 5 的结果可用于加强问题 4 中的不等式, 使之成为

$$E\left[\max_{0\leqslant k\leqslant n} |X_k|^\alpha\right] \leqslant \left(\frac{\alpha}{\alpha-1}\right)^\alpha E[\,|X_n|^\alpha\,].$$

试证明该不等式在 $\alpha = 2$ 的情形.

7. **群体消失.** 考虑生活在某一定环境的生物群体, 譬如说在地球上. 设 X_n 表示在时刻 n 活着的生物个体数, 这样 $\{0\}$ 是吸收状态, $X_n = 0$ 蕴涵着对于所有的 m 有 $X_{n+m} = 0$. 有理由假定对每个 N, 存在 $\delta > 0$, 使得如果 $X_n \leqslant N$, 则

$$\Pr[X_{n+1} = 0 | X_1, \cdots, X_n] \geqslant \delta,$$

$n = 1, 2, \cdots$ 令 $\mathcal{E}$ 表示群体终于消失的事件, 即

$$\mathcal{E} = \{X_k = 0, \text{对某个 } k = 1, 2, \cdots\}.$$

试证明以概率 1 或 $\mathcal{E}$ 出现或当 $n \to \infty$ 时 $X_n \to \infty$. 由于后者在一个有限的环境之中不可能出现, 因此, 最终群体必定消失.

8. 设 $Z, Y_0, Y_1, \cdots$ 是随机变量序列并且 $E[\,|Z|^2\,] < \infty$. 试证明 $X_n = E[Z|Y_0, \cdots, Y_n]$ 满足鞅均方收敛定理的条件.

9. 设 $\{X_n\}$ 是鞅, 且对于所有的 n 满足 $E[X_n^2] \leqslant K < \infty$. 若

$$\lim_{n\to\infty} \sup_{m\geqslant 1} |E[X_n X_{n+m}] - E[X_n]E[X_{n+m}]| = 0.$$

试证明 $X = \lim\limits_{n\to\infty} X_n$ 是一个常数, 即 X 是非随机的.

10. 设 $\{X_n\}$ 是鞅, 且对于所有的 n 满足 $E[X_n] = 0, E[X_n^2] < \infty$. 试证明

$$\Pr\left\{\max_{0\leqslant k\leqslant n} X_k > \lambda\right\} \leqslant \frac{E[X_n^2]}{E[X_n^2] + \lambda^2}, \qquad \lambda > 0.$$

提示：对每个 $c>0, (X_n+c)^2$ 是下鞅, 并对 $\lambda>0$ 应用最大值不等式得到

$$\Pr\left\{\max_{0\leqslant k\leqslant n} X_k>\lambda\right\}\leqslant \Pr\left\{\max_{0\leqslant k\leqslant n}(X_k+c)^2>(\lambda+c)^2\right\}$$
$$\leqslant \frac{E[(X_n+c)^2]}{(\lambda+c)^2},\qquad \text{对于所有的} c>0.$$

332 然后, 确定 c 值使得右端为最小.

11. 设 $\{X_n\}$ 是下鞅, 试证明

$$\lambda\Pr\left\{\max_{0\leqslant k\leqslant n} X_k<-\lambda\right\}\leqslant E[X_n^+]-E[X_0],\qquad \lambda>0.$$

12. 试证明：若 $\{X_n\}$ 是非负下鞅, 则

$$\lambda\Pr\left\{\max_{0\leqslant k\leqslant n} X_k\geqslant\lambda\right\}\leqslant E[X_0],\qquad \lambda>0.$$

(参见引理 5.2)

问题 13~16 均在如下条件之下考虑. 设 $\mathcal{B}_0=\{B_1,B_2,\cdots\}$ 是 Ω 的可数分割, 即 $\Omega=\bigcup_{n=1}^{\infty}B_n$, 且 $B_i\bigcap B_j=\varnothing(i\neq j)$. 令 $\mathcal{B}$ 是由 Ω 和所有 $\mathcal{B}_0$ 中集合的并构成的 σ 域, $\mathcal{B}_0$ 中集合的并可表为

$$B=\bigcup_{k=1}^{j}B_{n(k)}, 1\leqslant j\leqslant\infty,\quad \text{其中} B_{n(k)}\in\mathcal{B}_0.$$

13. 设 $\mathcal{B}$ 是由某个随机变量 Y 产生的 σ 域 (这样的随机变量 Y 至多具有可数个可能值). 试证明随机变量 X 是 $\mathcal{B}$ 可测的, 当且仅当存在某个实值函数 f 使得 $X=f(Y)$.

14. 设 X_1 和 X_2 是 $\mathcal{B}$ 可测随机变量. 试证明 $a_1X_1+a_2X_2$ 对任意实数 a_1,a_2 均为 $\mathcal{B}$ 可测的.

15. 设 Y 是 $\mathcal{B}$ 可测的, 且 $E[\,|Y|\,]<\infty$. 试证明, 若对于所有的有界非负 $\mathcal{B}$ 可测随机变量 $Z, E[YZ]\geqslant 0$, 则 $P[\{\omega:Y(\omega)\geqslant 0\}]=1$.

16. 试证明 X 是 $\mathcal{B}$ 可测的, 当且仅当存在某个实数列 $\{\alpha_k\}$ 使得

$$X(\omega)=\sum_{k=1}^{\infty}\alpha_k I_{B_k}(\omega),$$

其中

$$I_{B_j}(\omega)=\begin{cases}1, & \text{若 } \omega\in B_j,\\ 0, & \text{若 } \omega\notin B_j.\end{cases}$$

特别地, 请注意此时 $X(\omega)$ 在每个 B_j 上均为常数.

17. 固定 $\lambda > 0$. 设 $X_1, X_2, \cdots$ 是一随机变量序列且满足条件

$$E[\exp\{\lambda X_{n+1}\}|X_1, \cdots, X_n] \leqslant 1 \qquad \text{对于所有的} n.$$

令 $S_n = X_1 + \cdots + X_n \quad (S_0 = 0)$. 试证明

$$\Pr\left\{\sup_{n\leqslant 0}(x + S_n) > l\right\} \leqslant \mathrm{e}^{-\lambda(l-x)}, \quad \text{对于} x \leqslant l.$$

提示: 利用可选停止定理于非负上鞅 $\exp\{-\lambda(l - x - S_n)\}$.

333

18. 设 X 是一随机变量, 满足条件

$$\Pr\{-\varepsilon \leqslant X \leqslant +\varepsilon\} = 1, \tag{A}$$

和

$$E[X] \leqslant -\rho\varepsilon, \tag{B}$$

其中 $\varepsilon > 0$ 和 $\rho > 0$ 是已知的. 试证明对于 $\lambda = \varepsilon^{-1}\cdot \ \mathrm{In}[(1+\rho)/(1-\rho)]$,

$$E[\mathrm{e}^{\lambda X}] \leqslant 1,$$

并应用问题 17 的结果求下式的界

$$\Pr\left\{\sup_{n\geqslant 0}(x + S_n) > l\right\}, \quad \text{对} x < l,$$

其中 $S_n = X_1 + \cdots + X_n$, 且在已知 $X_1, X_2, \cdots, X_n$ 的条件下, X_{n+1} 的条件分布满足 (A) 和 (B).

19. 设 X 是满足如下条件的随机变量:

(a) $E[X] \leqslant m < 0$,

(b) $\Pr\{-1 \leqslant X \leqslant +1\} = 1$.

$X_1, X_2, \cdots$ 是一联合分布随机变量序列, 且在已知 $X_1, \cdots, X_n$ 的条件下, X_{n+1} 的条件分布满足 (a) 和 (b). 令 $S_n = X_1 + \cdots + X_n \quad (S_0 = 0)$ 且对于 $a < x$ 令

$$T_a = \min\{n : x + S_n \leqslant a\}.$$

试证明

$$E[T_a] \leqslant (1 + x - a)/|m|, \qquad a < x.$$

20. 设 $T, Y_0, Y_1, \cdots$ 是一列随机变量, T 的可能值为 $\{0, 1, \cdots\}$, 且对于每个 n, 事件 $\{T \geqslant n\}$ 由 $(Y_0, \cdots, Y_n)$ 所确定. 试问 T 关于 $\{Y_n\}$ 是否一定是马尔可夫时间? 若是请给出证明, 若不是请举出反例.

21. 设 $S_n=\xi_1+\cdots+\xi_n$, 其中 $\{\xi_k\}$ 是独立同分布正随机变量 ($\Pr\{\xi_k>0\}=1$). 试证明对任意的 $a>0$,

$$\sup_{n\geqslant 1} E\left[\frac{n}{a+S_n}\right]<\infty.$$

(ξ_k 假定仅取整数值的情况在 6.6 节例 (e) 已讨论过.)

22. 设 $Y_1,Y_2,\cdots$ 是独立随机变量, 且 $\Pr\{Y_k=+1\}=\Pr\{Y_k=-1\}=\frac{1}{2}$. 置 $S_k=Y_1+\cdots+Y_k$. 试证明

334

$$\Pr\{S_k<k \text{ 对所有 } k=1,\cdots,N|S_N=a\}=1-\frac{a}{N}.$$

23. 设 ξ_n 是非负随机变量且满足

$$E[\xi_{n+1}|\xi_1,\cdots,\xi_n]\leqslant\delta_n+\xi_n,$$

其中 $\delta_n>0$ 是常数且 $\Delta=\sum_{n=1}^{\infty}\delta_n<\infty$. 试证明当 $n\to\infty$ 时 ξ_n 以概率 1 收敛于一有限随机变量 ξ.

24. 在 $[0,1)$ 上的 Haar 函数定义如下:

$$H_1(t)\equiv 1,$$

$$H_2(t)=\begin{cases}1, & 0\leqslant t<\frac{1}{2},\\ -1, & \frac{1}{2}\leqslant t<1,\end{cases}$$

$$H_{2^n+1}(t)=\begin{cases}2^{n/2}, & 0\leqslant t<2^{-(n+1)},\\ -2^{n/2}, & 2^{-(n+1)}\leqslant t<2^{-n}, \qquad n=1,2,\cdots,\\ 0, & \text{其他},\end{cases}$$

$$H_{2^n+j}(t)=H_{2^n+1}\left(t-\frac{j-1}{2^n}\right),\qquad j=1,\cdots,2^n.$$

设 $f(z)$ 是 $[0,1]$ 上任一函数且满足

$$\int_0^1|f(z)|\mathrm{d}z<\infty.$$

定义 $a_k=\int_0^1 f(t)H_k(t)\mathrm{d}t$. 令 Z 是 $[0,1]$ 上的均匀分布. 试证明以概率 1

$$f(Z)=\lim_{n\to\infty}\sum_{k=1}^{n}a_kH_k(Z)$$

且

$$\lim_{n\to\infty}\int_0^1 |f(t)-\sum_{k=1}^{n} a_k H_k(t)|\mathrm{d}t = 0.$$

25. 设 $X_1, X_2, \cdots$ 是一列相互独立的随机变量且具有有限矩母函数 $\varphi_k(t) = E[\exp\{tX_k\}]$. 试证明, 若当 $n\to\infty$ 时有 $\Phi_n(t_0) = \prod_{k=1}^{n}\varphi_k(t_0) \to \Phi(t_0)$, 其中 $t_0 \neq 0, 0 < \Phi(t_0) < \infty$, 则 $S_n = X_1 + \cdots + X_n$ 以概率 1 收敛.

26. 设有一成功游程马尔可夫链, 0 是其吸收状态且具有转移概率 $P_{00} = 1$ 和 $P_{i,i+1} = p_i = 1 - P_{i,0},\ i = 1, 2, \cdots$. 若 $p_i \geqslant p_{i+1} \geqslant \cdots$, 令 a 是使得 $ap_{a-1}/(a-1) > 1 \geqslant (a+1)p_a/a$ 成立的唯一值. 定义

$$f(i) = \begin{cases} 0, & i = 0, \\ ap_i p_{i+1}\cdots p_{a-1}, & 1 \leqslant i < a, \\ i, & i \geqslant a. \end{cases}$$

335

(a) 证明对于所有的 $i = 0, 1, \cdots,\quad f(i) \geqslant i$.

(b) 证明对于所有的 i 有 $f(i) \geqslant E[f(X_{n+1})|X_n = i]$, 从而 $\{f(X_n)\}$ 是非负上鞅.

(c) 利用 (a) 和 (b) 验证对于所有的马尔可夫时间 T, $f(i) \geqslant E[X_T|X_0 = i]$.

(d) 证明 $f(i) = E[X_{T^*}|X_0 = i]$, 其中 $T^* = \min\{n \geqslant 0; X_n \geqslant a$ 或 $X_n = 0\}$. 这样, T^* 使得 $E[X_T|X_0 = i]$ 在所有的马尔可夫时间 T 中取最大值.

27. 设 $\Omega = \{\omega_1, \omega_2, \cdots\}$ 是一可数集, $\mathcal{F}$ 是由 Ω 的所有子集构成的 σ 域. 对于固定的 N, 令 $X_0, X_1, \cdots, X_N$ 是定义在 Ω 上一列随机变量, T 是关于 $\{X_n\}$ 的马尔可夫时间, 且满足 $0 \leqslant T \leqslant N$. 令 $\mathcal{F}_n$ 是由 $X_0, X_1, \cdots, X_n$ 生成的 σ 域, 定义 $\mathcal{F}_T$ 是由 $\mathcal{F}$ 中这样的集合 A 组成的, 它使得对任意 $n = 0, \cdots, N,\quad A\cap\{T = n\}$ 属于 $\mathcal{F}_n$. 即

$$\mathcal{F}_T = \{A : A \in \mathcal{F},\ \text{且} A\cap\{T = n\} \in \mathcal{F}_n,\ n = 0, \cdots, N\}.$$

试证明:

(a) $\mathcal{F}_T$ 是 σ 域,

(b) T 关于 $\mathcal{F}_T$ 可测,

(c) $\mathcal{F}_T$ 是由 $\{X_0, \cdots, X_T\}$ 生成的 σ 域, 其中 $\{X_0, \cdots, X_T\}$ 是定义在 Ω 上可变维数的向量值函数.

28. 设 $S_n = X_1 + \cdots + X_n$ 是零均值鞅, 其中对于所有的 $n, E[X_n^2] < \infty$. 试证明对任何单调实数列 $b_1 \leqslant \cdots \leqslant b_n \leqslant b_{n+1} \uparrow \infty$, 若 $\sum_{n=1}^{\infty} E[X_n^2]/b_n^2 < \infty$, 则以概率 1 $S_n/b_n \to 0$.

29. 设 X_n 是一保险公司在第 n 年末的总资产. 每年保险费收入为 $b>0$, 而第 n 年付出的赔偿费为 A_n, 所以 $X_{n+1}=X_n+b-A_n$. 假定 $A_1,A_2,\cdots$ 是相互独立的正态随机变量, 其均值 $\mu<b$, 方差为 σ^2. 若该公司的资产降为 0 以下即认为破产. 试证明

$$\Pr\{破产\}\leqslant \exp\{-2(b-\mu)X_0/\sigma^2\}.$$

30. 设 $Y_1,Y_2,\cdots$ 是具有有限均值 μ 的独立同分布正随机变量. 对于固定 $0<\beta<1$, 令 a 是使得 $u\geqslant \beta E[u\vee Y_1]=\beta E[\max\{u,Y_1\}]$ 成立的最小 u 值. 置 $f(x)=a\vee x$. 试证明 $\{\beta^n f(M_n)\}$ 是非负上鞅, 其中 $M_n=\max\{Y_1,\cdots,Y_n\}$, 而 a 对于所有的马尔可夫时间 T 有 $a=f(0)\geqslant E[\beta^T f(M_T)]$. 然后证明 $a=E[\beta^{T^*}M_{T^*}]$, 其中 $T^*=\min\{n\geqslant 1: Y_n\geqslant a\}$. 这样, T^* 使得 $E[\beta^T M_T]$ 在所有马尔可夫时间 T 中取最大值.

31. 设 $X,X_1,X_2,\cdots$ 是具有负均值 μ 和有限方差 σ^2 的独立同分布随机变量. 令 $S_0=0, S_n=X_1+\cdots+X_n$. 置 $M=\max\limits_{n\geqslant 0}S_n$, 由于 $\mu<0$, 我们知道 $M<\infty$. 设 $E[M]<\infty$.(事实上, 可以证明它是 $\sigma^2<\infty$ 的一个推论.) 定义 $r(x)=x^+=\max\{x,0\}$ 和 $f(x)=E[(x+M-E[M])^+]$.

336

(a) 证明对于所有的 $x, f(x)\geqslant r(x)$.

(b) 证明对于所有的 $x, f(x)\geqslant E[f(x+X)]$, 所以 $\{f(x+S_n)\}$ 是非负上鞅.

[提示: 验证并利用 M 和 $(X+M)^+$ 具有相同的分布这一事实.]

(c) 利用 (a) 和 (b) 证明, 对于所有的马尔可夫时间 $T, f(x)\geqslant E[(x+S_T)^+]$.[$(x+S_\infty)^+=\lim\limits_{n\to\infty}(X+S_n)^+=0$.]

32. (续上) 设 T^* 是马尔可夫时间.

$$T^*=\begin{cases}\min\{n\geqslant 0: x+S_n\geqslant E[M]\}, & 若 x+S_n\geqslant E[M] 对于某个 n,\\ \infty, & 若 x+S_n<E[M] 对于所有 n,\end{cases}$$

试证明, $f(x)=E[(x+S_{T^*})^+]$. 因此, T^* 使得 $E[r(S_T)]$ 在所有马尔可夫时间中取最大值.

33. 令 $\{X_n\}$ 是成功游动马尔可夫链, 具有转移概率 $P_{i,i+1}=p_i=1-P_{i,0},\ i=0,1,\cdots$. 设 $0<p_i<1$ 且 $p_i\geqslant p_{i+1}\geqslant\cdots$. 固定 $0<\beta<1$, a 是使得 $a\beta p_{a-1}/(a-1)>1\geqslant(a+1)\beta p_a/a$ 成立的唯一值. 定义

$$f(i)=\begin{cases}a\beta^{a-i}p_i\cdot p_{i+1}\cdots p_{a-1}, & i<a,\\ i, & i\geqslant a.\end{cases}$$

(a) 证明 $f(i)\geqslant i$ 对于所有 i 成立.

(b) 证明 $f(i)\geqslant \beta E[f(X_n)|X_{n-1}=i]$, 所以 $\{\beta^n f(X_n)\}$ 是非负上鞅.

(c) 利用 (a) 和 (b) 验证对于所有的马尔可夫时间 $T, f(i)\geqslant E[\beta^T X_T|X_0=i]$.

(d) 证明 $f(i)=E[\beta^{T^*}X_{T^*}|X_0=i]$, 其中 $T^*=\min\{n\geqslant 0:\ X_n\geqslant a\}$. 这样, T^* 使得 $E[\beta^T X_T|X_0=i]$ 在所有的马尔可夫时间 T 中取最大值.

34. 设 Z_n 是具有转移概率矩阵 $P(i,j)$ 的马尔可夫链. 令 $f(i)$ 是一有界函数, 并对于所有的 i 定义 $F(i)=\sum_j P(i,j)f(j)-f(i)$. 试验明以概率 1,

$$\frac{F(Z_1)+\cdots+F(Z_n)}{n}\to 0,\qquad n\to\infty.$$

提示: 利用问题 28 的结果.

35. 设 $\{X_n\}$ 是一鞅, 且满足条件 $\sup_n E[\,|X_n|\,]<\infty$. 试推导表示式 $X_n=X_n^{(1)}-X_n^{(2)}$, 其中 $\{X_n^{(i)}\}(i=1,2)$ 是具有有界均值的非负鞅.

提示: $Z_n^N=E[\,|X_{N+1}|Y_0,\cdots,Y_n\,]$ 关于 N 递增, 所以 $Z_n=\lim_{N\to\infty}Z_n^N$ 存在且是非负的, 并且 $E[\,|Z_n|\,]\leqslant\sup_n E[\,|X_n|\,]<\infty$. 试证明 $\{Z_n\}$ 是一个鞅, 然后令 $X_n^{(1)}=Z_n,\ \ X_n^{(2)}=Z_n-X_n$.

36. 设 $\{X_n\}$ 是一具有有限均值的下鞅, 且 $X_0=0$. 试推导表示式 $X_n=X_n'+X_n''$, 其中 $\{X_n'\}$ 是鞅而 $X_n''\leqslant X_{n+1}''$ 是不减过程.

提示: 见初等问题 20.

337

37. 设 $\{X_n\}$ 是鞅, 且 $Y=\sup_n|X_{n+1}-X_n|$ 具有有限均值. 令 A_1 是 $\{X_n\}$ 收敛的事件, A_2 是 $\limsup X_n=+\infty$ 和 $\liminf X_n=-\infty$ 的事件. 试证明 $\Pr\{A_1\}+\Pr\{A_2\}=1$. 换句话说, $\{X_n\}$ 或者收敛, 或者振动于正负无穷大之间.

提示: 对每个 $k,\widetilde{X}_n=X_{T\wedge n}$ 收敛, 其中 $T=\min\{n:X_n\geqslant k\}$, 因为 $\widetilde{X}_n\leqslant k+Y$, 所以 $\sup_n E[\widetilde{X}_n]<\infty$. 这样, 如果对于每个 k, $\limsup X_n>k$, 则 $\{X_n\}$ 不收敛. 同样分析可应用于 $\{-X_n\}$.

38. 设 $\varphi(\xi)$ 是对称函数, 关于 $|\xi|$ 是不减的, $\varphi(0)=0$, 且 $\{\varphi(Y_j)\}_{j=0}^n$ 是下鞅. 固定 $0=u_0\leqslant u_1\leqslant\cdots\leqslant u_n$. 试证明

$$\Pr\{|Y_j|\leqslant u_j;1\leqslant j\leqslant n\}\geqslant 1-\sum_{j=1}^n\frac{E[\varphi(Y_j)]-E[\varphi(Y_{j-1})]}{\varphi(u_j)}.$$

(若 $\varphi(\xi)=\xi^2,u_1=\cdots=u_n=\lambda$, 我们得到科尔莫戈罗夫不等式.)

提示: 若 $\{|Y_j|\leqslant u_j\}$, 令 I_j 为 1; 其他情况 I_j 为 0, 则

$$\begin{aligned}\Pr\{|Y_j|\leqslant u_j;1\leqslant j\leqslant n\}&=E\left[\prod_{j=1}^n I_j\right]\\&\geqslant E\left[\prod_{j=1}^{n-1}I_j\left(1-\frac{\varphi(Y_n)}{\varphi(u_n)}\right)\right]\end{aligned}$$

$$\geqslant E\left[\prod_{j=1}^{n-1} I_j\left(1-\frac{\varphi(Y_{n-1})}{\varphi(u_{n-1})}\right)\right]-\frac{E[\varphi(Y_n)]-E[\varphi(Y_{n-1})]}{\varphi(u_n)},$$

然后连续利用 $\{\varphi(Y_n)\}$ 是下鞅及 $\{u_n\}$ 的递增性. 重复之.

39. 设 $\{Y_n\}$ 是非负下鞅. b_n 是不增的正数序列, 且满足 $\sum_{n=1}^{\infty}(b_n-b_{n+1})E[Y_n]<\infty$. 试证明

$$\lambda\Pr\{\sup_{k\geqslant 1} b_k Y_k>\lambda\}<\sum_{k=1}^{\infty}(b_k-b_{k+1})E[Y_k].$$

40. 设 $\{X_n\}$ 是一族随机变量, 令 $\varphi(\xi)$ 是定义在 $\xi>0$ 上的正函数, 且满足

$$\frac{\varphi(\xi)}{\xi}\to\infty,\qquad \text{当}\xi\to\infty.$$

若

$$\sup_{m\geqslant 1} E[\varphi(|X_m|)]\leqslant K<\infty,$$

338 试证明 $\{X_n\}$ 是一致可积的.

41. 设 $R_k(x)$ 是 Rademacher 函数 $P_k(x)=\text{sign}\sin(2^{k+1}\pi x)$. 定义

$$L_n(x)=\prod_{k=1}^{n}(1+a_k R_k(x)),\qquad 0\leqslant x\leqslant 1,$$

其中 a_k 是常数.

令 $\mathcal{F}_n, n=0,1,\cdots$ 是由单位区间上的分割

$$\left(\frac{k}{2^{n+1}},\frac{k+1}{2^{n+1}}\right),\qquad k=0,1,\cdots,2^{n+1}$$

生成的 σ 域, $\mu(\mathrm{d}x)$ 是概率测度, 它赋予 $\mathcal{F}_n$ 的每个基本分割区间的概率为 $1/2^{n+1}$. 试证明

(a) L_n 是 $\mathcal{F}_n$ 可测的.

(b) L_n 是适应于 $\mathcal{F}_n$ 的鞅.

附 记

Doob[1] 创立了鞅理论并证明这个理论具有广泛的应用.

参考书目

[1] J.L. Doob, *Stochastic Processes*. Wiley, New York, 1953.

[2] Paul-André Meyer, *Martingales and Stochastic Integrals.* SpringerVerlag, Berlin, New York, 1972(Lecture Notes in Mathematics No. 284).

[3] J.Neveu, *Mathematical Foundations of the Calculus of Probability.* Holden-Day, San Francisco, 1965.

[4] J.Neveu, *Martingales à Temps Discret.* Masson, Paris, 1972.

339

第 7 章　布朗运动

R.Brown 于 1827 年观察到悬浮在液体中的小粒子呈现出不停地无规则运动. 本章的布朗运动过程在历史上是在试图解释这个现象的过程中而发展起来的. 如今, 布朗运动过程一般化及其推广大量出现在纯粹和应用科学的诸多领域中, 如经济, 通讯理论, 生物, 管理科学和数理统计.

本章前 4 节是关于这个方面内容的导引, 随后一节是利用鞅的方法计算若干个与布朗运动相联系的期望和概率. 后者要求 6.5 节作为其预备知识. 最后一节处理比较专门的论题.

7.1　背景材料

某些特别类型的随机过程已被广泛深入地研究过. 布朗运动正是最著名的而且在历史上首先比较彻底研究过的一个过程. 我们仅仅提出它的显著特点的一小部分, 希望由此能促使读者进一步学习这个过程详尽的讨论.

布朗运动作为物理现象首先被英国植物学家布朗于 1827 年发现. 这个现象的数学描述由爱因斯坦于 1905 年由物理定律导出. 此后, 这个课题得到了巨大的发展. 其物理理论被 Smoluchowski, Fokker, Planck, Burger, Furth, Ornstein, Uhlenbeck, Chandrasekhar, Kramers 等人进一步完善. 然而数学上的描述却发展比较慢, 因为准确地数学描述这个模型非常困难, 相比之下, 物理学家根据这个模型可凭直觉简单地找到一些问题的答案. 许多结果 Bachelier 于 1900 年在他的博士论文中虽已推测到, 但该理论的简明数学公式首先是 Wiener 在 1918 年的博士论文以及后来的文章中提出来的 (见本章末参考书目).

340

利用随机过程的术语讲, 布朗运动是连续时间, 连续状态空间的马尔可夫过程的一个例子.

用 $X(t)$ 表示布朗运动的一个粒子在 x 轴方向的分量 (作为时间 t 的函数). 设 X_0 是粒子在时刻 t_0 时的位置, 即 $X(t_0) = x_0$. 令 $p(x, t|x_0)$ 表示在已知 $X(t_0) = x_0$ 之下 $X(t + t_0)$ 的条件概率密度. 我们假定转移所服从的概率规律关于时间是平稳的, 因此 $p(x, t|x_0)$ 不依赖于初始时刻 t_0.

由于 $p(x, t|x_0)$ 是 x 的密度函数, 我们有如下性质

$$p(x,t|x_0) \geqslant 0, \quad \int_{-\infty}^{\infty} p(x,t|x_0)\mathrm{d}x = 1. \tag{1.1}$$

进一步, 我们规定, 对很小的 $t, X(t+t_0)$ 很可能靠近 $X(t_0)=x_0$. 形式上我们要求

$$\lim_{t\to 0} p(x,t|x_0)=0, \quad \text{对于} x\neq x_0. \tag{1.2}$$

爱因斯坦根据物理学定律指出 $p(x,t|x_0)$ 一定满足偏微分方程

$$\frac{\partial p}{\partial t}=D\frac{\partial^2 p}{\partial x^2} \tag{1.3}$$

这就是所谓的扩散方程, 其中 D 是扩散系数. 悬浮在液体或气体中的小粒子呈现布朗运动是由于它们受到液体或气体分子碰撞而引起的. D 的值由公式 $D=2RT/Nf$ 计算, 其中 R 是气体常数, T 是温度, N 是 Avogadro 数, f 是磨擦系数. 选择适当的尺度, 我们可取 $D=1/2$. 我们可直接验证

$$p(x,t|x_0)=\frac{1}{\sqrt{2\pi t}}\exp\left(-\frac{1}{2t}(x-x_0)^2\right) \tag{1.4}$$

是 (1.3) 的解, 实际上在边界条件 (1.2) 及 (1.1) 之下解是唯一的. (关于 (1.3) 解的唯一性问题需要准确地系统地阐述, 其分析要求非常细微, 这已超出本书范围.) 341

导出 (1.3) 的另一途径是用离散随机游动来逼近. 考虑在整数集上的对称随机游动 (见 2.2 节中的例 B). 令 $p_k(n)$ 表示在这种随机游动中从原点出发在时刻 n 到达原点右边第 k 步的概率. 这个过程的切普曼 — 科尔莫戈罗夫关系式 (第 2 章公式 (3.2)) 是

$$p_k(n+1)=\frac{1}{2}p_{k+1}(n)+\frac{1}{2}p_{k-1}(n),$$

由此有

$$p_k(n+1)-p_k(n)=\frac{1}{2}[p_{k+1}(n)-2p_k(n)+p_{k-1}(n)]. \tag{1.5}$$

上式左边是时间导数的离散变形, 而右边部分是关于空间变量二阶导数的离散变形的一半. 通过适当的取极限过程, 其中两次转移之间的时间缩小到 0, 同时转移步数的大小也压缩到 0, 我们就可以从 (1.5) 过渡到 (1.3).

具体地, 设两次转移之间的时间长度是 Δ, 并且每步的长度为 η, 则 (1.5) 类似变为

$$\frac{p_{k\eta}((n+1)\Delta)-p_{k\eta}(n\Delta)}{\Delta}=\frac{\frac{1}{2}[p_{(k+1)\eta}(n\Delta)-2p_{k\eta}(n\Delta)+p_{(k-1)\eta}(n\Delta)]}{\Delta} \tag{1.6}$$

现在令 Δ 和 η 趋向于 0, 保持关系式 $\Delta=\eta^2$, 并且同时令 n 和 k 趋于 ∞ 使得 $k\eta\to x$ 而 $n\Delta\to t$. 则 $p_{k\eta}(n\Delta)\to p(x,t|0)$, 而 (1.6) 形式上变为 (1.3).

我们并不想把上述步骤严格化, 显然概念上是简单的, 但要使得它精确化尚需比较细致的分析.

关于 $p_k(n)$ 的另一个类型的极限过程需要用到中心极限定理. 我们记

$$p_k(n) = \Pr\{X_1 + X_2 + \cdots + X_n = k\},$$

其中 $\{X_i\}$ 表示掷一个均匀硬币的一系列结果 (即 $X_i = 1$ 表示正面, $X_i = -1$ 表示反面, 每种情况出现的概率为 1/2). 由中心极限定理 (见 1.1 节) 得

$$\lim_{n\to\infty} \sum_{k=-\infty}^{\sqrt{n}x} p_k(n) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-u^2/2)\mathrm{d}u. \tag{1.7}$$

(1.6) 的极限关系式和 (1.7) 的极限关系式本质上是相同的, 它们都与 "随机过程的不变原理" 相联系. 这些问题可以精确化, 但已超出本书范围.

342

7.2 布朗运动的联合概率

转移概率密度函数 (1.4) 仅给出 $X(t) - X(0)$ 的概率分布. 对布朗运动的完整的描述由下面定义来完成:

定义 2.1 布朗运动是具有下面性质的随机过程 $\{X(t); t \geqslant 0\}$:

(a) 每个增量 $X(t+s) - X(s)$ 服从均值为 0、方差为 $\sigma^2 t$ 的正态分布, 其中 σ 是固定参数.

(b) 对于任意两个不相交的时间区间 $[t_1, t_2]$, $[t_3, t_4]$, 不妨设 $t_1 < t_2 \leqslant t_3 < t_4$, 增量 $X(t_4) - X(t_3)$ 和 $X(t_2) - X(t_1)$ 是服从 (a) 中分布的相互独立的随机变量, 对于任意 n 个不相交的时间区间也有类似结论.

(c) $X(0) = 0$ 和 $X(t)$ 在 $t = 0$ 连续.

上述假设意味着位移 $X(t+s) - X(s)$ 与过去相独立, 或者, 如果我们知道 $X(s) = x_0$, 则对于 $\tau < s$, $X(\tau)$ 的取值并不影响 $X(t+s) - X(s)$ 的概率规律. 用式子表示就是, 如果 $t > t_0 > t_1 > \cdots > t_n$, 我们有

$$\begin{aligned} &\Pr[X(t) \leqslant x | X(t_0) = x_0, X(t_1) = x_1, \cdots, X(t_n) = x_n] \\ &= \Pr[X(t) \leqslant x | X(t_0) = x_0]. \end{aligned} \tag{2.1}$$

这说明了布朗运动过程的马尔可夫性. 然而, 我们应当指出, 独立增量的假设 (b) 实际上比马尔可夫限制性更强.

在条件 $X(0) = 0$ 之下, $X(t)$ 的方差是 $\sigma^2 t$, 因此 σ^2 有时称为过程的方差参数. 过程 $\widetilde{X}(t) = X(t)/\sigma$ 是具有方差参数为 1 的布朗运动, 称为**标准布朗运动**. 因此, 我们总是可以把任意一个布朗运动简化为标准布朗运动. 随后的大部分结果我们仅对后者进行推导.

由定义的 (a), 当 $\sigma^2=1$ 时, 我们有

$$\Pr[X(t)\leqslant x|X(t_0)=x_0]=\Pr[X(t)-X(t_0)\leqslant x-x_0]$$
$$=\frac{1}{\sqrt{2\pi(t-t_0)}}\int_{-\infty}^{x-x_0}\exp\left[-\frac{\alpha^2}{2(t-t_0)}\right]\mathrm{d}\alpha. \quad (2.2)$$

343

定义中 (b) 和 (a) 的相容性可由著名的正态分布的性质推得. 例如, 如果 $t_1\leqslant t_2\leqslant t_3$, 则

$$X(t_3)-X(t_1)=[X(t_3)-X(t_2)]+[X(t_2)-X(t_1)].$$

在右边我们有均值为 0、方差分别为 t_3-t_2 及 t_2-t_1 的独立正态随机变量, 因此它们之和是均值为 0、方差为 t_3-t_1 的正态随机变量, 这正是所要求的.

在条件 $X(0)=0$ 之下, 利用 (2.1) 和 (2.2) 导出 $X(t_1),X(t_2),\cdots,X(t_n)(t_1<t_2<\cdots<t_n)$ 的联合密度是不困难的. 实际上, 我们只要知道 $X_1=X(t_1)=x_1,X_2-X_1=x_2-x_1,\cdots$, 直至 $X_n-X_{n-1}=x_n-x_{n-1}$ 的概率密度, 由定义的 (b) 我们马上得到密度函数的以下表达式:

$$f(x_1,\cdots,x_n)=p(x_1,t_1)p(x_2-x_1,t_2-t_1)\cdots p(x_n-x_{n-1},t_n-t_{n-1}), \quad (2.3)$$

其中

$$p(x,t)=\frac{1}{\sqrt{2\pi t}}\exp(-x^2/2t). \quad (2.4)$$

利用公式 (2.3), 原则上我们可以计算任何想要的条件概率.

根据马尔可夫性我们知道, 如果 $t_1<t_2<t_3$, 在已知 $X(t_1)$ 和 $X(t_2)$ 条件下, $X(t_3)$ 的条件密度和仅仅已知 $X(t_2)$ 条件下 $X(t_3)$ 的条件密度是相同的.

然而, 已知 $X(t_1)$ 和 $X(t_3)$ 的条件下, 求 $X(t_2)$ 的密度也是有趣的. 为了确定起见, 我们假设 $X(t_1)=X(t_3)=0$, 具体一点, 比如说, $t_1=0,t_3=1$, 且 $t_2=t(0<t<1)$.

由 (2.3), $X(t)$ 和 $X(1)$ 的联合密度是

$$f(x,y)=\frac{1}{2\pi\sqrt{t(1-t)}}\exp\left[-\frac{1}{2}\left\{\frac{x^2}{t}+\frac{(y-x)^2}{1-t}\right\}\right].$$

由此推得在已知 $X(0)=X(1)=0$ 条件之下 $X(t)$ 的条件密度 $f_t(x|X(0)=X(1)=0)$ 是

$$\frac{1}{\sqrt{2\pi t(1-t)}}\exp\left[-\frac{1}{2}\frac{x^2}{t(1-t)}\right],\quad -\infty<x<\infty.$$

特别地有 $E_c(X(t))=0$ 和 $E_c(X^2(t))=t(1-t)$, 其中 E_c 为在条件 $X(0)=X(1)=0$ 之下的期望. 用相同的方法可得到更加一般的结果.

344

定理 2.1 在已知 $X(t_1)=A$ 和 $X(t_2)=B$ 的条件下, 对于 $t_1<t<t_2$, $X(t)$ 的条件密度服从正态分布, 其均值和方差分别为

$$A+\frac{B-A}{t_2-t_1}(t-t_1) \text{ 和 } \frac{(t_2-t)(t-t_1)}{t_2-t_1}.$$

这可以用如下方法归结为前面的情况. 前面已指出, 在条件 $X(t_1)=A$ 和 $X(t_2)=B$ 之下的随机变量 $X(t)$, 与在条件 $X(0)=0$, $X(t_2-t_1)=B-A$ 之下的随机变量 $A+X(t-t_1)$ 具有相同的密度. 显然, 这与在条件 $X(0)=0$, $X(t_2-t_1)=0$ 之下的随机变量

$$A+X(t-t_1)+\frac{(t-t_1)}{t_2-t_1}(B-A)$$

具有相同的密度.

7.3 轨道的连续性和最大值变量

布朗运动过程的物理学起源提示我们, 它的任一可能的实现 $X(t)$, 作为粒子由于介质分子连续碰撞引起运动的 x 坐标的图形, 是一连续函数. 这个事实是正确的, 但它的严格证明要求比较细致的分析, 我们将在 7.7 节中叙述.

样本轨道 $X(t)$ 虽然连续却非常奇怪, 它的导数处处不存在. 这个事实也是很深入的. 关于布朗运动轨道构造的一个比较完整叙述可以在 P.Levy 和 Mckean 的著作中找到 (见本章末参考书目).

利用轨道的连续性我们将说明怎样推演出有关布朗运动各种有趣的概率表达式. 下面的计算阐述了所谓**反射原理**的应用.

记住 $X(t)$ 的连续性, 现在我们考虑样本轨道的集合 $X(t):0\leqslant t\leqslant T, X(0)=0$, 且具有性质 $X(T)>a(a>0)$. 由于 $X(t)$ 是连续的, $X(0)=0$, 因此存在一个时刻 τ(这是一个随机变量, 取决于具体的样本轨道), 它是 $X(t)$ 首次达到值 a 的时刻.

对于 $t>\tau$, 我们作 $X(t)$ 关于 $X=a$ 的反射, 得到

345

$$\widetilde{X}(t)=\begin{cases} X(t), & \text{若 } t<\tau, \\ a-[X(t)-a], & \text{若 } t>\tau, \end{cases}$$

(见图 7-1). 注意到由于 $X(T)>a$, 故 $\widetilde{X}(T)<a$. 因为当已知 $X(\tau)=a$ 时, 对于 $t>\tau$ 时轨道的概率规律关于值 $x>a$ 和 $x<a$ 是对称的, 并且独立于在时刻 τ 前的历史, 从而反射论证法把每条具有 $X(T)>a$ 的样本轨道展现为以相同概率出现的两条样本轨道 $X(t)$ 和 $\widetilde{X}(t)$, 使得

$$\max_{0\leqslant u\leqslant T} X(u)\geqslant a \text{ 和 } \max_{0\leqslant u\leqslant T} \widetilde{X}(u)\geqslant a.$$

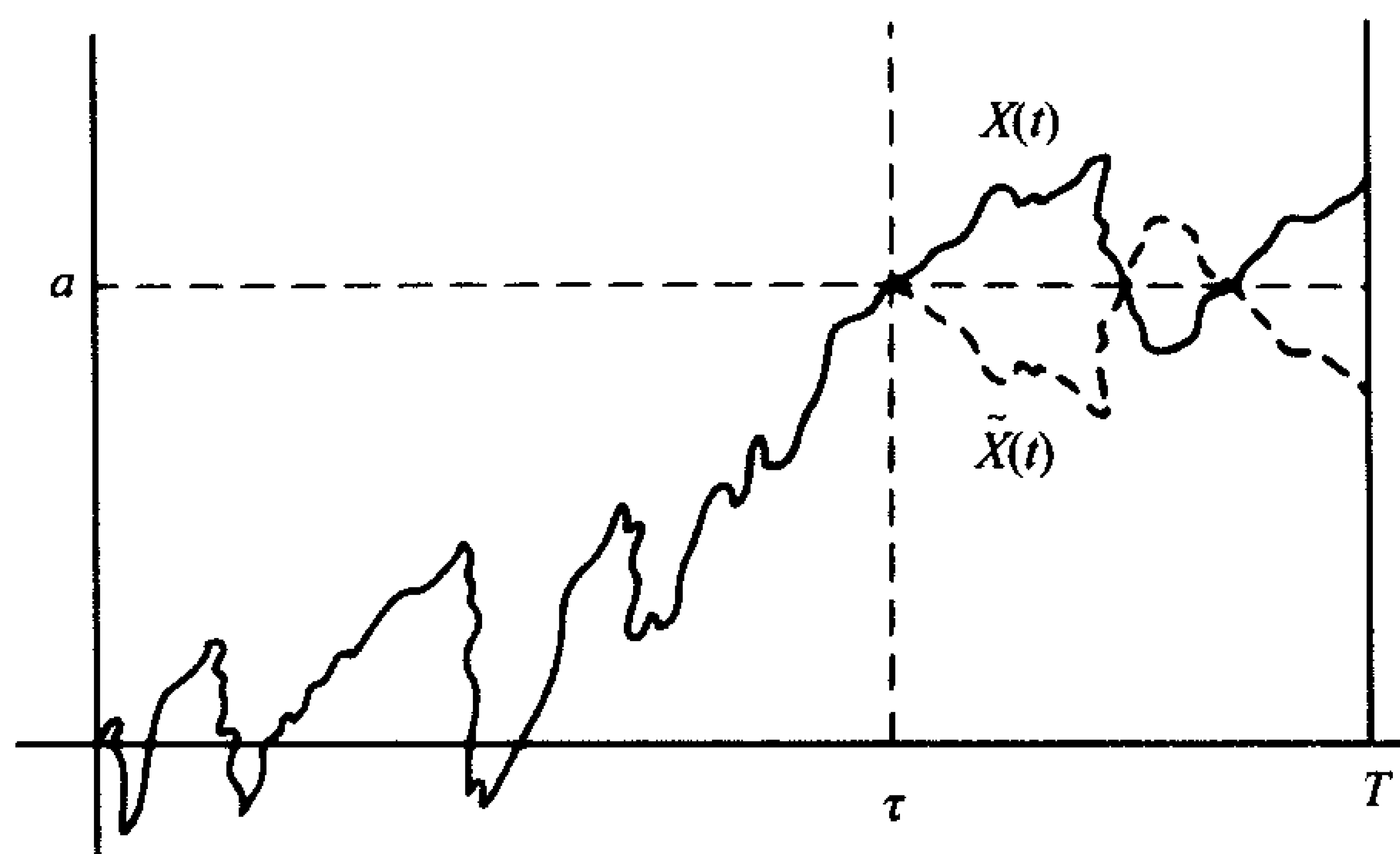


图 7-1

反之, 根据这个对应的性质, 每个样本函数 $X(t)$ 当 $\max\limits_{0\leqslant u\leqslant T} X(u)\geqslant a$ 时, 可从具有相同概率的两个样本函数 $X(t)$ 中的任何一个得到, 其中一条除 $X(T)=a$ 之外必有 $X(T)>a$, 但 $\Pr\{X(T)=a\}=0$. 于是, 我们可断定在条件 $X(0)=0$ 之下有

$$\Pr\{\max_{0\leqslant u\leqslant T} X(u)\geqslant a\}=2\Pr\{X(T)>a\}=\frac{2}{\sqrt{2\pi T}}\int_a^\infty \exp(-x^2/2T)\mathrm{d}x. \tag{3.1}$$

上面的论证法不能认为是完整的, 虽然这个方法是典型的, 它是研究具有连续轨道马尔可夫过程的基础. (这样的过程称为扩散过程.) 严格的论述必须使用关于马尔可夫时间 (见 14.4 节) 的强马尔可夫性, 其中马尔可夫时间是对应于首次从 0 到达 a 的事件.

346

借助于 (3.1), 我们可以确定在条件 $X(0)=0$ 之下, 到达 $a>0$ 的首达时间分布. 设 T_a 表示 $X(t)$ 首达 a 的时间, 其中 $X(0)=0$. 则显然有

$$\Pr\{T_a\leqslant t\}=\Pr\{\max_{0\leqslant u\leqslant t} X(u)\geqslant a|X(0)=0\}. \tag{3.2}$$

但依照 (3.1).

$$\Pr\{\max_{0\leqslant u\leqslant t} X(u)\geqslant a|X(0)=0\}=2\Pr\{X(t)>a\}=\frac{2}{\sqrt{2\pi t}}\int_a^\infty \exp\left[-\frac{1}{2}\frac{x^2}{t}\right]\mathrm{d}x$$

所以

$$\Pr\{T_a\leqslant t\}=\sqrt{\frac{2}{\pi t}}\int_a^\infty \exp\left[-\frac{1}{2}\frac{x^2}{t}\right]\mathrm{d}x,$$

作变换 $x=y\sqrt{t}$, 得

$$\Pr\{T_a\leqslant t\}=\sqrt{\frac{2}{\pi}}\int_{a/\sqrt{t}}^\infty \exp\left[-\frac{y^2}{2}\right]\mathrm{d}y. \tag{3.3}$$

随机变量 T_a 的密度函数可由关于 t 微分 (3.3) 得到. 这样

$$f_{T_a}(t|X(0)=0)\mathrm{d}t = \frac{a}{\sqrt{2\pi}} t^{-3/2} \exp\left[-\frac{a^2}{2t}\right] \mathrm{d}t. \tag{3.4}$$

由于布朗运动的对称性和位置齐次性, 对于分布 (3.4) 我们推得

$$\begin{aligned}
&\Pr\{\min_{0\leqslant u\leqslant t} X(u) \leqslant 0|X(0)=a\} \\
&= \Pr\{\max_{0\leqslant u\leqslant t} X(u) \geqslant 0|X(0)=-a\} \qquad (\text{由对称性}) \\
&= \Pr\{\max_{0\leqslant u\leqslant t} X(u) \geqslant a|X(0)=0\} = \Pr\{T_a \leqslant t\} \qquad (\text{由齐次性}) \\
&= \frac{a}{\sqrt{2\pi}} \int_0^t u^{-3/2} \exp\left[-\frac{a^2}{2u}\right] \mathrm{d}u, \quad a>0. \qquad (3.5)
\end{aligned}$$

可用另一方法表达 (3.5): 如果 $X(t_0)=a$, 则 $X(t)$ 在 t_0 和 t_1 之间至少有一个零点的概率 $P(a)$ 是

347

$$P(a) = \frac{|a|}{\sqrt{2\pi}} \int_0^{t_1-t_0} \exp\left[-\frac{a^2}{2u}\right] u^{-3/2} \mathrm{d}u. \tag{3.6}$$

利用此公式我们可以计算如果 $X(0)=0$, 则 $X(t)$ 在区间 (t_0,t_1) 至少有一次为 0 的概率 α.

事实上, 我们可限制 $X(t_0)$ 的可能值. 如果 $X(t_0)=a$, 则 $X(t)$ 在区间 (t_0,t_1) 至少有一次为 0 的概率是 $P(a)$. 由全概率公式得

$$\alpha = \int_0^\infty P(a)\Pr\{|X(t_0)|=a|X(0)=0\}\mathrm{d}a = \sqrt{\frac{2}{\pi t_0}} \int_0^\infty P(a) \exp\left[-\frac{a^2}{2t_0}\right] \mathrm{d}a. \tag{3.7}$$

把 (3.6) 代入, 然后改变积分顺序得到

$$\begin{aligned}
\alpha &= \sqrt{\frac{2}{\pi t_0}} \int_0^\infty \exp\left[-\frac{a^2}{2t_0}\right] \frac{a}{\sqrt{2\pi}} \left(\int_0^{t_1-t_0} \exp\left[-\frac{a^2}{2u}\right] u^{-3/2}\mathrm{d}u\right) \mathrm{d}a \\
&= \frac{1}{\pi\sqrt{t_0}} \int_0^{t_1-t_0} u^{-3/2} \left(\int_0^\infty a \exp\left[-\frac{a^2}{2}\left(\frac{1}{u}+\frac{1}{t_0}\right)\right] \mathrm{d}a\right) \mathrm{d}u. \qquad (3.8)
\end{aligned}$$

里面的积分可准确地计算, 经化简之后, 我们得到

$$\alpha = \frac{\sqrt{t_0}}{\pi} \int_0^{t_1-t_0} \frac{\mathrm{d}u}{(t_0+u)\sqrt{u}}.$$

由变量替换 $u=t_0v^2$, 得

$$\alpha = \frac{2}{\pi} \int_0^{\sqrt{(t_1-t_0)/t_0}} \frac{\mathrm{d}v}{1+v^2} = \frac{2}{\pi}\arctan\sqrt{\frac{t_1-t_0}{t_0}},$$

利用标准的三角关系公式, 我们可写为

$$\sqrt{\frac{t_0}{t_1}} = \cos\frac{\pi\alpha}{2}, \text{ 或 } \alpha = \frac{2}{\pi}\arccos\sqrt{\frac{t_0}{t_1}}.$$

综上所述, 我们有

定理 3.1 在已知 $X(0)=0$ 的条件之下, $X(t)$ 在区间 (t_0,t_1) 至少有一个零点的概率是

$$\alpha = \frac{2}{\pi}\arccos\sqrt{\frac{t_0}{t_1}}.$$

借助同样的"反射原理", 我们来解决下面的问题: 对 $x>0$ 和 $y>0$, 求

$$A_t(x,y) = \Pr\{X(t)>y, \min_{0\leqslant u\leqslant t} X(u)>0|X(0)=x\}. \tag{3.9}$$

348

为此, 我们从如下的明显关系式开始

$$\begin{aligned}&\Pr\{X(t)>y|X(0)=x\}\\&=A_t(x,y)+\Pr\{X(t)>y, \min_{0\leqslant u\leqslant t} X(u)\leqslant 0|X(0)=x\}.\end{aligned} \tag{3.10}$$

反射原理应用于最后一项.

图 7-2 是引导我们分析的一个适当图形. 我们可以推出

$$\begin{aligned}&\Pr\{X(t)>y, \min_{0\leqslant u\leqslant t} X(u)\leqslant 0|X(0)=x\}\\&=\Pr\{X(t)<-y, \min_{0\leqslant u\leqslant t} X(u)\leqslant 0|X(0)=x\}\\&=\Pr\{X(t)<-y|X(0)=x\}.\end{aligned} \tag{3.11}$$

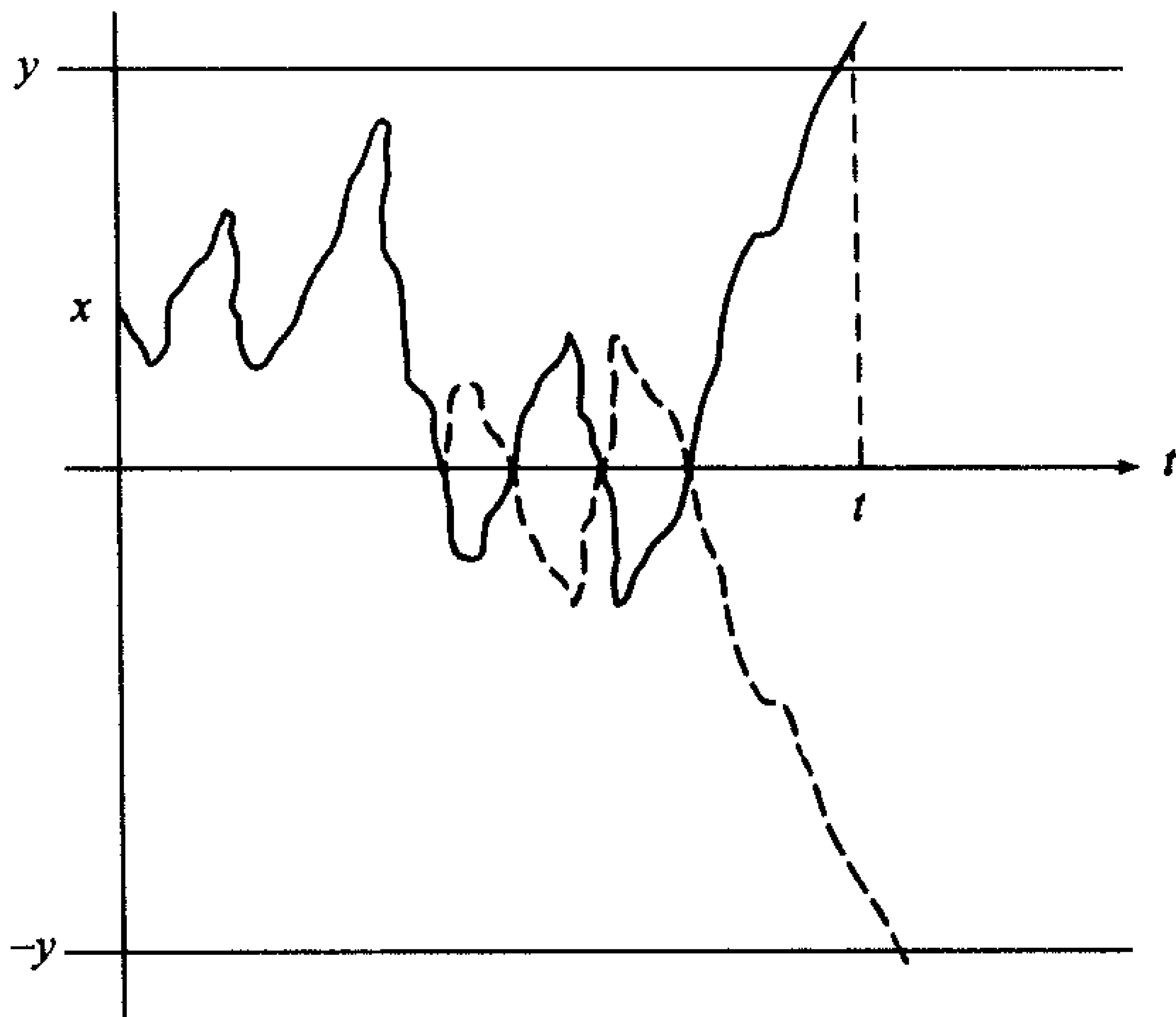


图 7-2

(3.11) 成立的理由如下：考虑一条从 $x > 0$ 出发, 满足 $X(t) > y$, 并在某个中间时间 τ 到达 0 的轨道. 对这条轨道在时间 τ 之后关于 0. 实施反射, 我们得到一条从 x 出发并且在时刻 t 到达比 $-y$ 更小的值的轨道. 这蕴涵着 (3.11) 前面两项相等. 最后的两项等式显然可从它们意义得到, 因为条件 $\min_{0 \leqslant u \leqslant t} X(u) \leqslant 0$ 对于 $X(t) < -y(y > 0)$ 是不必要的. 把 (3.11) 代入 (3.10) 得到

349

$$\begin{aligned}
A_t(x, y) &= \Pr\{X(t) > y | X(0) = x\} - \Pr\{X(t) < -y | X(0) = x\} \\
&= \Pr\{X(t) > y - x | X(0) = 0\} \\
&\quad - \Pr\{X(t) < -(y + x) | X(0) = 0\} \quad (\text{由位置齐次性}) \\
&= \Pr\{X(t) > y - x | X(0) = 0\} - \Pr\{X(t) > y + x | X(0) = 0\} \quad (\text{由对称性}) \\
&= \int_{y-x}^{y+x} p(u, t) \mathrm{d}u, \qquad (3.12)
\end{aligned}$$

其中 $p(u, t) = (2\pi t)^{-1/2} \exp\{-u^2/2t\}$ 是布朗运动的转移概率密度函数.

作为反射原理的最后一个应用, 我们导出

$$M(t) = \max_{0 \leqslant u \leqslant t} X(u) \quad 和 \quad Y(t) = M(t) - X(t)$$

的联合概率密度函数.

由反射原理推得 (见图 7-3)

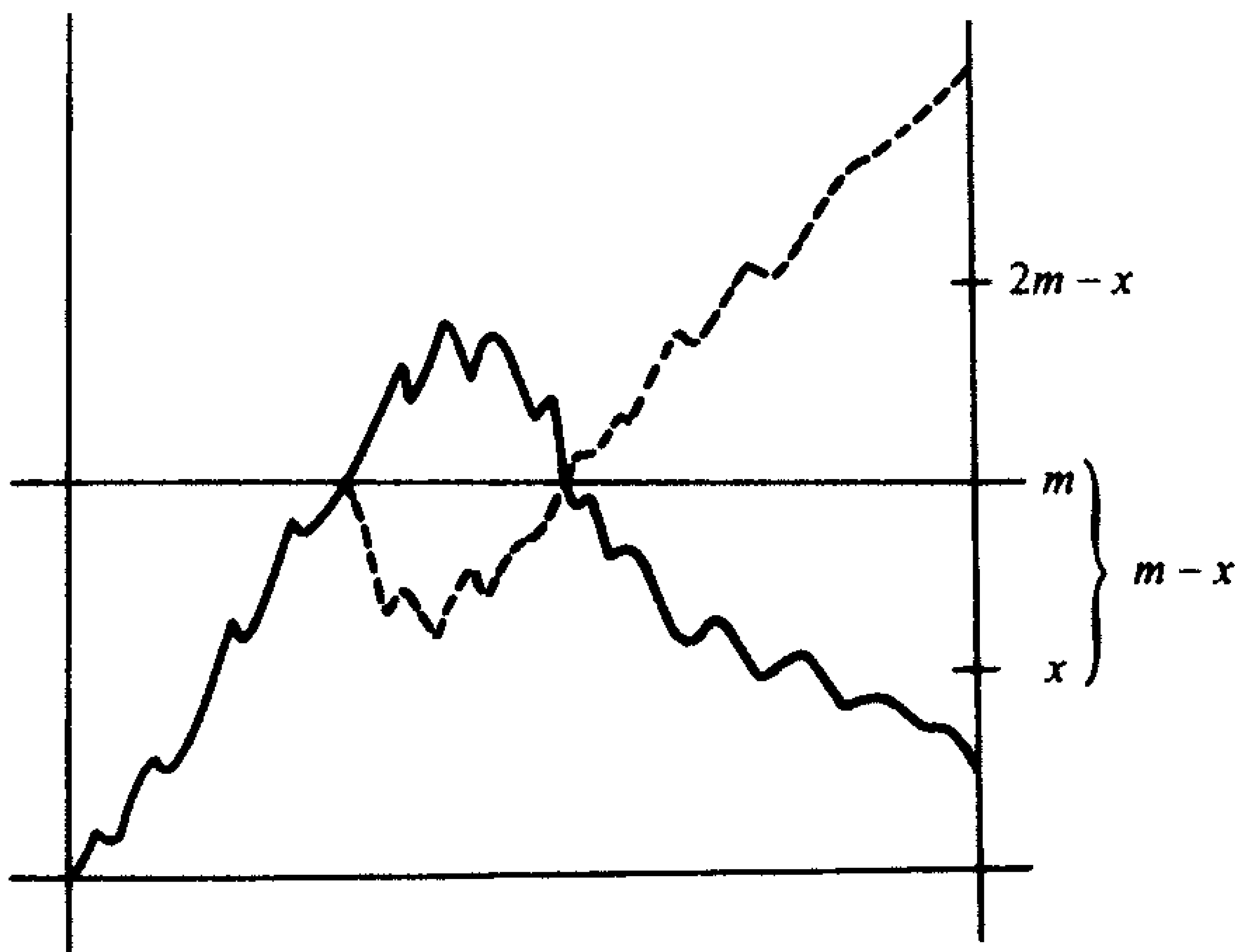


图 7-3

$$\begin{aligned}
\Pr\{M(t) \geqslant m, X(t) \leqslant x\} &= \Pr\{X(t) \geqslant 2m - x\} \\
&= 1 - \Phi\left(\frac{2m - x}{\sqrt{t}}\right), \quad m \geqslant 0, m \geqslant x,
\end{aligned}$$

其中 $\Phi(x)$ 是标准正态密度 $\phi(x)=(1/\sqrt{2\pi})\exp(-x^2/2)$ 的分布函数. 关于 x 微分, 然后关于 m 微分, 改变符号, 得到 $M(t)$ 和 $X(t)$ 的联合密度函数.

计算式子如下

$$-\frac{\mathrm{d}}{\mathrm{d}m}\frac{\mathrm{d}}{\mathrm{d}x}\left\{1-\Phi\left(\frac{2m-x}{\sqrt{t}}\right)\right\}=-\frac{\mathrm{d}}{\mathrm{d}m}\left\{\frac{1}{\sqrt{t}}\phi\left(\frac{2m-x}{\sqrt{t}}\right)\right\}$$
$$=\frac{2m-x}{t}\frac{2}{\sqrt{t}}\phi\left(\frac{2m-x}{\sqrt{t}}\right),$$

这里利用了初等关系式

$$\frac{\mathrm{d}}{\mathrm{d}x}\phi(x)=-x\phi(x).$$

350

用 $f(m,x)$ 表示这个联合密度, 显然我们有

$$f(m,x)=\sqrt{\frac{2}{\pi t^3}}(2m-x)\exp[-(2m-x)^2/2t],\quad\left\{\begin{array}{l}0\leqslant m,\\ x\leqslant m.\end{array}\right.\tag{3.13}$$

为了得到 $M(t)$, $Y(t)=M(t)-X(t)$ 的联合密度 $g(m,y)$, 我们有

$$\Pr\{M(t)\leqslant a,Y(t)\leqslant b\}=\int_{m\leqslant a}\int_{y\leqslant b}f(m,m-y)\mathrm{d}y\mathrm{d}m.$$

因此所求的联合密度是 $g(m,y)=f(m,m-y)$ 或

$$g(m,y)=\sqrt{\frac{2}{\pi t^3}}(m+y)\exp[-(m+y)^2/2t],\quad m\geqslant 0,y\geqslant 0.\tag{3.14}$$

7.4 变形和推广

我们将阐述, 如果 $X(t)$ 是标准布朗运动, 则过程

$$X_1(t)=cX(t/c^2),\quad \text{对于固定的 } c>0,$$

$$X_2(t)=\left\{\begin{array}{ll}tX(1/t), & t>0,\\ 0, & t=0,\end{array}\right.$$

及

$$X_3(t)=X(t+h)-X(h),\quad \text{对于固定的 } h>0.$$

每一个都是标准布朗运动的变形.

351

我们验证定义 2.1 的要求. 显然, 在这些过程中任一增量 $X_i(t+s)-X_i(t)$ 是零均值的正态分布, 并且在不相交的时间区间上的增量显然是相互独立的随机变量. 为

了继续验证这些是布朗运动, 我们需要求每种情况下的方差 $E[\{X_i(s+t)-X_i(s)\}^2]$. 有

$$
\begin{aligned}
E[\{X_1(t+s)-X_1(s)\}^2] &= c^2E[\{X((t+s)/c^2)-X(s/c^2)\}^2] \\
&= c^2\{(t+s)/c^2 - s/c^2\} = t, \\
E[\{X_2(t+s)-X_2(s)\}^2] &= E\left[\left\{sX\left(\frac{1}{s}\right)-(t+s)X\left(\frac{1}{t+s}\right)\right\}^2\right] \\
&= s^2E\left[\left\{X\left(\frac{1}{s}\right)-X\left(\frac{1}{t+s}\right)\right\}^2\right]+t^2E\left[\left\{X\left(\frac{1}{t+s}\right)\right\}^2\right] \\
&= s^2\left\{\frac{1}{s}-\frac{1}{t+s}\right\}+t^2\frac{1}{t+s} \\
&= t,
\end{aligned}
$$

以及

$$
E[\{X_3(t+s)-X_3(s)\}^2] = E[\{X(t+h+s)-X(h+s)\}^2] = t.
$$

为完成此分析, 需验证每个 $X_i(t)$ 在原点是连续的. 这个性质对于 $X_1(t)$ 和 $X_3(t)$ 是显然的, 但对于 $X_2(t)$ 需要某些论证. 对后者只须证明如下

$$
\Pr\left\{\lim_{t\to\infty}\frac{X(t)}{t}=0|X(0)=0\right\}=1.
$$

(见问题 14 和问题 15.)

布朗运动其他的变形导出新的重要的随机过程. 下面是一些例子

A. 在原点反射的布朗运动

令 $\{X(t), t\geqslant 0\}$ 是布朗运动. 与随机过程

$$
\begin{aligned}
Y(t) &= |X(t)|, \quad t\geqslant 0 \\
&= \begin{cases} X(t), & \text{如果 } X(t)\geqslant 0, \\ -X(t), & \text{如果 } X(t)<0, \end{cases}
\end{aligned}
$$

352 具有相同分布的随机过程称为在原点反射的布朗运动, 我们简称为**反射布朗运动**.

因为布朗运动的位置对称性, 所以反射布朗运动是马尔可夫的. 的确, 在标准布朗运动情况下 ($\sigma^2=1$),

$$
\begin{aligned}
&\Pr\{Y(t_n+s)\leqslant z|Y(t_0)=x_0,\cdots,Y(t_n)=x_n\} \\
&\quad=\Pr\{-z\leqslant X(t_n+s)\leqslant +z|X(t_0)=\pm x_0,\cdots,X(t_n)=\pm x_n\}
\end{aligned}
$$

$$
\begin{aligned}
&= \Pr\{-z \leqslant X(t_n+s) \leqslant +z | X(t_0)=x_0,\cdots,X(t_n)=x_n\} \quad (\text{由对称性}) \\
&= \Pr\{-z \leqslant X(t_n+s) \leqslant +z | X(t_n)=x_n\} \\
&= \int_{-z}^{+z} p(y-x_n,s)\mathrm{d}y, \qquad (4.1)
\end{aligned}
$$

其中

$$
p(x,t)=\frac{1}{\sqrt{2\pi t}}\exp\{-x^2/2t\}, \qquad (4.2)
$$

并且 $0 \leqslant t_0 < t_1 < \cdots < t_n, s > 0$. 这样反射布朗运动提供了另外一个连续时间马尔可夫过程的例子, 它的样本轨道 $Y(t)$ 是连续的. 状态空间是由非负实数组成的.

由于 $Y(t)$ 的矩和 $|X(t)|$ 的矩是相同的, 反射布朗运动的均值和方差容易计算. 例如, 在条件 $Y(0)=0$ 之下, 有

$$
\begin{aligned}
E[Y(t)] &= \int_{-\infty}^{+\infty} |x| p(x,t)\mathrm{d}x \\
&= 2\int_0^{\infty} \frac{x}{\sqrt{2\pi t}}\exp\{-x^2/2t\}\mathrm{d}x \\
&= \sqrt{\frac{2t}{\pi}},
\end{aligned}
$$

其中, 我们通过变量替换 $y=x/\sqrt{t}$ 求积分值. 另外,

$$
\begin{aligned}
Y(t)\text{ 的方差} &= E[Y(t)^2]-E[Y(t)]^2 \\
&= E[|X(t)|^2]-2t/\pi \\
&= (1-2/\pi)t.
\end{aligned}
$$

353

在 (4.1) 中进行变量替换, 易知对于 $t>0$, $Y(t)$ 是连续型随机变量, 具有如下转移概率密度函数

$$
p_t(x,y)=p(y-x,t)+p(y+x,t), \qquad (4.3)
$$

且

$$
\Pr\{\alpha \leqslant Y(t) \leqslant \beta | Y(0)=x\} = \int_{\alpha}^{\beta} p_t(x,y)\mathrm{d}y,
$$

其中 $p(x,t)$ 由 (4.2) 给定.

B. 被原点吸收的布朗运动

假设布朗运动的初始值 $X(0)=x$ 是正的 [1], 令 τ 是过程首次到达 0 的时间. 与

1. 为定义在 $X(0)=x$ 条件下的布朗运动, 或"从 x 开始的布朗运动"只要在定义 2.1 的条款 (c) 中将"$X(0)=0$"改为"$X(0)=x$".

随机过程

$$Z(t)=\begin{cases} X(t), & \text{对于} \quad t\leqslant\tau, \\ 0, & \text{对于} \quad t>\tau, \end{cases} \tag{4.4}$$

具有相同分布的随机过程称为被原点吸收的布朗运动, 简称为**被吸收的布朗运动.**

被吸收的布朗运动也是连续时间马尔可夫过程. 我们验证如下形式的马尔可夫性:

$$\begin{aligned} &\Pr\{Z(t_n+s)>y|Z(t_0)=x_0,\cdots,Z(t_{n-1})=x_{n-1},Z(t_n)=x\} \\ &\quad=\Pr\{Z(t_n+s)>z|Z(t_n)=x\} \end{aligned}$$

其中 $x>0$ 和 $0<t_0<\cdots<t_n, s>0$. (当 $x=0$ 时比较容易, 留给读者验证.) 如 (4.1) 那样, 我们通过与 $Z(t)$ 相联系的布朗运动过程 $X(t)$ 计算这些概率. 由条件 $x>0$ 推得

$$\min_{0\leqslant u\leqslant t_n} X(u)>0.$$

因此,

$$\begin{aligned} &\Pr\{Z(t_n+t)>y|Z(t_0)=x_0,\cdots,Z(t_{n-1})=x_{n-1},Z(t_n)=x\} \\ &\quad=\Pr\{Z(t_n+t)>y|Z(t_0)=x_0,\cdots,Z(t_{n-1})=x_{n-1},Z(t_n)=x, \\ &\qquad\min_{0\leqslant u\leqslant t_n} X(u)>0\} \\ &\quad=\Pr\{X(t_n+t)>y,\min_{0\leqslant u\leqslant t} X(t_n+u)>0|X(t_0)=x_0,\cdots, \\ &\qquad X(t_{n-1})=x_{n-1},X(t_n)=x\} \\ &\quad=\Pr\{X(t)>y,\min_{0\leqslant u\leqslant t} X(u)>0|X(0)=x\} \\ &\quad=A_t(x,y), \end{aligned}$$

354

其中 $A_t(x,y)$ 由 (3.9) 定义, 并且由 (3.12) 计算, 故

$$\begin{aligned} A_t(x,y)&=\int_y^{\infty}[p(u-x,t)-p(u+x,t)]\mathrm{d}u \\ &=\int_y^{y+2x} p(u-x,t)\mathrm{d}u. \end{aligned}$$

在条件 $Z(0)=x>0$ 之下, $Z(t)$ 是随机变量, 其分布兼有离散和连续的部分. 离散部分是

$$\begin{aligned} &\Pr\{Z(t)=0|Z(0)=x\}=1-A_t(x,0) \\ &\quad=1-\int_0^{2x} p(u-x,t)\mathrm{d}u \\ &\quad=1-\int_{-x}^{x} p(u,t)\mathrm{d}u \end{aligned}$$

$$= 2\int_x^{\infty} p(u,t)\mathrm{d}u$$
$$= 2\int_0^{\infty} p(u+x,t)\mathrm{d}u.$$

在区域 $z>0, Z(t)$ 是一连续型随机变量, 并且对于 $0<a<b$,

$$\Pr\{a<Z(t)<b|Z(0)=x\} = A_t(x,a)-A_t(x,b)$$
$$= \int_a^b [p(u-x,t)-p(u+x,t)]\mathrm{d}u.$$

这样, 被吸收的布朗运动连续部分的转移概率密度函数是

$$p_t(x,y)=p(y-x,t)-p(y+x,t).$$

C. 漂移布朗运动

设 $\{\widetilde{X}(t), t\geqslant 0\}$ 是布朗运动. 漂移布朗运动是与

$$X(t)=\widetilde{X}(t)+\mu t, \quad t\geqslant 0,$$

有相同分布的随机过程, 其中 μ 是常数, 称为**漂移参数**. 我们也可以用平行于定义 2.1 的方式来描述漂移布朗运动.

355

定义 4.1 漂移布朗运动是具有下面性质的随机过程 $\{X(t); t\geqslant 0\}$:

(a) 每个增量 $X(t+s)-X(s)$ 是均值为 μt、方差为 $\sigma^2 t$ 的正态分布; μ,σ 是固定的常数.

(b) 对于任何两个不相交的时间区间 $[t_1,t_2],[t_3,t_4], t_1<t_2\leqslant t_3<t_4$, 增量 $X(t_4)-X(t_3)$ 和 $X(t_2)-X(t_1)$ 是与 (a) 中所给定的分布相同的独立随机变量. 对于任意 n 个不相交的时间区间也有类似性质.

(c) $X(0)=0$ 并且 $X(t)$ 在 $t=0$ 连续.

如前, 由定义推知位移 $X(t+s)-X(s)$ 与过去独立, 换句话说, 如果我们知道 $X(s)=x_0$, 对于 $\tau<s, X(t)$ 的值并不影响 $X(t+s)-X(s)$ 的条件概率规律. 写成式子就是, 如果 $t>t_0>t_1>\cdots>t_n$,

$$\Pr\{X(t)\leqslant x|X(t_0)=x_0, X(t_1)=x_1,\cdots,X(t_n)=x_n\}$$
$$=\Pr\{X(t)\leqslant x|X(t_0)=x_0\}. \tag{4.5}$$

这就说明了过程的马尔可夫性. 然而, 我们强调独立增量假设 (b) 实际上比马尔可夫性更强. 利用定义条款 (a), 我们有

$$\Pr\{X(t)\leqslant x|X(t_0)=x_0\}=\Pr\{X(t)-X(t_0)\leqslant x-x_0\}$$

$$= \int_{-\infty}^{x-x_0} \frac{1}{\sqrt{2\pi(t-t_0)}\sigma} \exp\left\{-\frac{(y-\mu(t-t_0))^2}{2(t-t_0)\sigma^2}\right\} dy$$
$$= \int_{-\infty}^{\{x-x_0-\mu(t-t_0)\}|\sigma} p(t-t_0, y) dy,$$

其中

$$p(t,x) = \frac{1}{\sqrt{2\pi t}} \exp\{-x^2/2t\}.$$

与 $\mu \neq 0$ 时, 过程不再是对称的, 反射论证法不能用于计算过程最大值的分布. 356 我们下一节将利用鞅理论来计算这个分布.

D. 几何布朗运动

设 $\{X(t), t \geqslant 0\}$ 是带有漂移参数 μ 和扩散系数 σ^2 的布朗运动. 由

$$Y(t) = e^{X(t)}, \quad t \geqslant 0,$$

所定义的过程有时称为**几何布朗运动**. 其状态空间是区间 $(0, \infty)$.

由于 $Y(t) = Y(0)e^{X(t)-X(0)}$, 利用正态分布的特征函数, 我们计算

$$E[Y(t)|Y(0)=y] = yE[e^{X(t)-X(0)}]$$
$$= y\exp\left\{t\left(\mu + \frac{1}{2}\sigma^2\right)\right\},$$

和

$$E[Y(t)^2|Y(0)=y] = y^2 E[e^{2[X(t)-X(0)]}]$$
$$= y^2 \exp\left\{t\left[2\mu + \frac{1}{2}4\sigma^2\right]\right\},$$

故 $Y(t)$ 的方差是

$$\begin{aligned}\mathrm{Var}[Y(t)|Y(0)=y] &= E[Y(t)^2|Y(0)=y] - \{E[Y(t)|Y(0)=y]\}^2 \\ &= y^2\left\{\exp\left[2t\left(\mu+\frac{1}{2}2\sigma^2\right)\right] - \exp\left[2t\left(\mu+\frac{1}{2}\sigma^2\right)\right]\right\} \\ &= y^2\exp\left[2t\left(\mu+\frac{1}{2}\sigma^2\right)\right][\exp(t\sigma^2)-1].\end{aligned}$$

7.5 用鞅方法计算若干布朗运动的量

将可选抽样定理应用于与布朗运动相关的鞅中可以迅速计算许多重要的量. 我们在第一章曾指出标准布朗运动 $\{X(t); t \geqslant 0\}$ 是鞅, 但这绝不意味着我们感兴趣的只是这个鞅. 过程

$$U(t) = X^2(t) - t$$

和

$$V(t)=\exp\left\{\lambda X(t)-\frac{1}{2}\lambda^2 t\right\} \tag{5.1}$$

关于标准布朗运动 $\{X(t)\}$ 都是鞅, 其中 λ 是任意实常数. 我们直接予以证明. 对于 $0\leqslant t_1\leqslant\cdots\leqslant t_n=t$ 和 $s>0$, 我们有 357

$$\begin{aligned}E[U(t+s)|X(t_1),\cdots,X(t_n)]&=E[X^2(t+s)|X(t)]-(t+s)\\&=E[\{X(t+s)-X(t)\}^2|X(t)]\\&\quad+2E[X(t)\{X(t+s)-X(t)\}|X(t)]\\&\quad+E[X^2(t)|X(t)]-(t+s)\\&=s+2\times 0+X^2(t)-(t+s)\\&=U(t).\end{aligned}$$

类似地,

$$\begin{aligned}E[V(t+s)|X(t_1),\cdots,X(t_n)]&=V(t)\times E\left[\exp\{\lambda[X(t+s)-X(t)]-\frac{1}{2}\lambda^2 s\}\right]\\&=V(t).\end{aligned}$$

附注 把例 (5.1) 的鞅置于一个更加容易得到的结构中是有好处的. 设马尔可夫过程 $\{X(t);t\geqslant 0\}$ 具有平稳转移概率密度

$$p(t,x|y)\mathrm{d}x=\Pr\{x\leqslant X(t)<x+\mathrm{d}x|X(0)=y\},$$

与其相联系的鞅的一般构造可叙述如下. 设 $u(x,t)$ 是函数方程

$$u(x,s)=\int p(t,\xi|x)u(\xi,t+s)\mathrm{d}\xi,\qquad s,t>0. \tag{5.2}$$

的解, 则 $Z(t)=u(X(t),t)$ 是关于由 $X(t)$ 的直至时刻 t 为止的历史产生的 σ 域族 $\mathcal{F}_t=\sigma\{X(u);0\leqslant u<t\}$ 的鞅, 倘若 $E[|Z(t)|]<\infty$.

证明 我们计算

$$\begin{aligned}E[Z(t+s)|\mathcal{F}(s)]&=E[Z(t+s)|X(s)]\quad(\text{由马尔可夫性})\\&=E[u(X(t+s),t+s)|X(s)]\\&=\int p(t,\xi|X(s))u(\xi,t+s)\mathrm{d}\xi\\&=u(X(s),s)\quad(\text{利用函数方程})\\&=Z(s).\end{aligned}$$

这就证明了 $\{Z(t)\}$ 是一个鞅.

由直接计算表明

358

$$u(x,t)=x^2-t \quad 和 \quad u(x,t)=\exp\left\{\lambda x-\frac{1}{2}\lambda^2 t\right\}$$

满足函数方程 (5.2), 此时应取

$$p(t,x|y)=\frac{1}{\sqrt{2\pi t}}\exp\{-(x-y)^2/2t\}.$$

在布朗运动的情况下, 可以进一步证明, 如果 $u(x,t)$ 是充分可微分的, 并且满足关系式 (5.2), 则 $u(x,t)$ 还满足

$$\frac{\partial u}{\partial t}+\frac{1}{2}\frac{\partial^2 u}{\partial x^2}=0. \tag{5.3}$$

在第 15 章关于扩散过程中, 我们将提出对最后这个结论的一个比较完整的证明. 反之也是正确的, 当 (5.3) 成立时也可推得 (5.2).

下面是应用鞅 (5.1) 进行计算的例子. 设 $a<0<b$ 是已知的, 令 $T=T_{ab}$ 是过程到达 a 或 b 的首达时间:

$$T_{ab}=\inf\{t\geqslant 0: X(t)=a \text{ 或 } X(t)=b\},$$

记 $T\wedge n=\min\{T,n\}$.

由于 $\{U(t)\}$ 是鞅, 则 $E[U(T\wedge n)]=E[U(0)]=0$, 由此得到

$$E[T\wedge n]=E[X^2(T\wedge n)]\leqslant(|a|+b)^2.$$

这样

$$E[T]=\lim_{n\to\infty}\int_0^n \Pr\{T>t\}\mathrm{d}t=\lim_{n\to\infty}E[T\wedge n]\leqslant(|a|+b)^2.$$

这里重要在于 $T=T_{ab}$ 是有限的, 甚至 T_{ab} 具有有限均值.

设 u 是过程 $\{X(t)\}$ 到达 a 之前到达 b 的概率, 即

$$u=\Pr\{X(T_{ab})=b\}.$$

由于 $\{X(t)\}$ 是鞅, $T=T_{ab}$ 是有限的, 并且 $\{X(t\wedge T)\}$ 是有界的. 这样, 应用可选停止定理有

$$\begin{aligned}0&=E[X(T)]\\&=a\Pr\{X(T)=a\}+b\Pr\{X(T)=b\}\\&=a[1-u]+bu,\end{aligned}$$

所以

$$u = \Pr\{X(T_{ab}) = b\} = \frac{|a|}{|a| + b}.$$

359

现在我们回到鞅 $\{U(t)\}$, 注意到 $E[U(T)] = 0$, 计算得

$$\begin{aligned} E[T_{ab}] &= E[X^2(T_{ab})] \\ &= a^2[1-u] + b^2 u \\ &= a^2 \frac{b}{|a|+b} + b^2 \frac{|a|}{|a|+b} \\ &= |a|b. \end{aligned}$$

通过改变已给过程的原点和尺度, 我们可以计算具有方差参数 σ^2 和初始位置 $X(0) = x$ 的布朗运动的相应量. 其结果如下:

定理 5.1 设 $\{X(t); t > 0\}$ 是具有方差 σ^2 和 $X(0) = x$ 的布朗运动. 给定 a, b 满足 $a < x < b$, 设 T 是过程首达 a 或 b 的时间, 则

$$\Pr\{X(T) = b | X(0) = x\} = (x - a)/(b - a),$$

并且

$$E[T | X(0) = x] = \frac{1}{\sigma^2}(b - x)(x - a).$$

下面我们转向研究带有漂移参数 $\mu \neq 0$ 和方差 σ^2 的布朗运动 $\{X(t), t \geqslant 0\}$. 对每一实数 λ,

$$V(t) = \exp\left\{\lambda X(t) - \left(\lambda\mu + \frac{1}{2}\lambda^2\sigma^2\right)t\right\} \tag{5.4}$$

定义一个鞅. 我们选取

$$\lambda_0 = -2\mu/\sigma^2,$$

所以 (5.4) 指数的第二项为 0. 则 $V_0(t) = \exp\{\lambda_0 X(t)\}$ 是一个鞅. T_{ab} 是过程到达 $a < 0$ 或 $b > 0$ 的首达时间, 应用可选停止定理, 得

$$\begin{aligned} 1 &= E[V_0(T_{ab})] \\ &= \Pr\{X(T_{ab}) = a\}\exp\{\lambda_0 a\} + \Pr\{X(T_{ab}) = b\}\exp\{\lambda_0 b\}, \end{aligned}$$

并且

$$\Pr\{X(T_{ab}) = b\} = \frac{1 - \exp\{\lambda_0 a\}}{\exp\{\lambda_0 b\} - \exp\{\lambda_0 a\}},$$

其中 $\lambda_0 = -2\mu/\sigma^2$. 再者, 我们可移动原点以处理 $X(0) = x$ 的情况.

360

定理 5.2 设 $\{X(t); t \geqslant 0\}$ 是带有漂移参数 $\mu \neq 0$ 和方差 σ^2 的布朗运动, 假设 $X(0) = x$. 则过程在到达 $a < x$ 之前到达 $b > x$ 的概率由下式给出:

$$\Pr\{X(T_{ab}) = b | X(0) = x\} = \frac{\exp(-2\mu x/\sigma^2) - \exp(-2\mu a/\sigma^2)}{\exp(-2\mu b/\sigma^2) - \exp(-2\mu a/\sigma^2)}.$$

推论 5.1 设 $X(t)$ 是带有漂移参数 $\mu < 0$ 的布朗运动. 设

$$W = \max_{0 \leqslant t < \infty} X(t) - X(0).$$[1]

则 W 具有指数分布

$$\Pr\{W \geqslant w\} = e^{-\lambda w}, \quad w \geqslant 0$$

其中 $\lambda = 2|\mu|/\sigma^2$.

证明 在定理 5.2 的公式中我们令 $a \to -\infty$. 由于 $\mu < 0, \exp(-\mu a/\sigma^2) \to 0$. 这样,

$$\lim_{a \to -\infty} \Pr\{X(T_{ab}) = b | X(0) = x\} = \exp[2\mu(b - x)/\sigma^2].$$

但当 $a \to -\infty$ 时, 左边部分是过程到达 b 的概率, 或过程最大值超过 b 的概率. 因此, 对于 $w = b - x$,

$$\begin{aligned}\Pr\{W \geqslant w\} &= \Pr\{\max_{0 \leqslant t < \infty} X(t) > b | X(0) = x\} \\ &= e^{-\lambda w},\end{aligned}$$

其中 $\lambda = 2|\mu|/\sigma^2$, 证毕 ■

作为最后一个例子, 我们来计算首次通过单壁的时间的拉普拉斯变换. 我们令 $z > 0$ 是固定的, 并且 $T = T_z$ 是过程首次到达水平 z 的时间:

$$T = T_z = \begin{cases} \inf\{t : X(t) \geqslant z\}, & \text{如果对某个 } t \geqslant 0 \text{ 有 } X(t) \geqslant z, \\ \infty, & \text{如果对所有 } t \geqslant 0 \text{ 有 } X(t) < z. \end{cases}$$

设置 $\theta = \lambda\mu + \frac{1}{2}\lambda^2\sigma^2$. 则 $V(t) = \exp\{\lambda X(t) - \theta t\}$ 是鞅, 并且, 若 $X(0) = 0$, 则

$$1 = E[V(T \wedge t)],$$

或

361

$$1 = E[\exp\{\lambda X(T \wedge t) - \theta(T \wedge t)\}].$$

1. 若 $X(t)$ 在 $[0, \infty)$ 无最大值, 则取上确界代替这里的 $\max_{0 \leqslant t < \infty} X(t)$.—— 译者注

现设 $\lambda \geqslant 0$ 充分可以保证 $\theta \geqslant 0$, 则

$$0 \leqslant V(T \wedge t) \leqslant e^{\lambda z},$$

从而, 利用第 6 章引理 3.3, 可以取极限, 得

$$\lim_{t\to\infty} V(t \wedge T) = \begin{cases} \exp\{\lambda z - \theta T\}, & \text{如果 } T < \infty, \\ 0, & \text{如果 } T = \infty. \end{cases}$$

于是

$$1 = \lim_{t\to\infty} E[V(T \wedge t)] = e^{\lambda z} E[e^{-\theta T}]$$

或

$$E[e^{-\theta T}] = e^{-\lambda z}.$$

接着讨论 θ 与 λ 的关系. 我们有

$$\frac{1}{2}\sigma^2\lambda^2 + \mu\lambda - \theta = 0,$$

或

$$\lambda = \frac{-\mu \pm \sqrt{\mu^2 + 2\sigma^2\theta}}{\sigma^2}.$$

我们要求 $\lambda \geqslant 0$, 因此

$$\lambda = \frac{1}{\sigma^2}(\sqrt{\mu^2 + 2\sigma^2\theta} - \mu).$$

当 $\mu < 0, T$ 是有缺欠的随机变量, 即 T 以正概率取无限, 并且

$$\begin{aligned} \Pr\{T < \infty\} &= \lim_{\theta\to 0} E[e^{-\theta T}] = \lim_{\theta\to 0} \exp\left[-\frac{z}{\sigma^2}(\sqrt{\mu^2 + 2\sigma^2\theta} - \mu)\right] \\ &= \exp(-2z|\mu|/\sigma^2), \end{aligned}$$

这和推论 5.1 是一致的. 当 $\mu \geqslant 0$ 时, 以概率 1 有 $T < \infty$, 并且拉普拉斯变换是

$$E[e^{-\theta T}] = \exp\left[-\frac{z}{\sigma^2}(\sqrt{\mu^2 + 2\sigma^2\theta} - \mu)\right]. \tag{5.5}$$

362

对于带有漂移参数 $\mu \geqslant 0$ 的布朗运动, 可以求变换 (5.5) 的逆, 并得到 T_z 的概率密度函数的显示表达式. 我们叙述此结果如下.

定理 5.3 设 $X(t)$ 是带有漂移参数 $\mu \geqslant 0$ 的布朗运动, 且 $z > X(0) = x$ 是已知的. 令 T_z 是过程首次到达水平 z 的时间. 则在 $X(0) = x$ 条件之下, T_z 具有概率密度函数

$$f(t; x, z) = \frac{(z - x)}{\sigma\sqrt{2\pi t^3}} \exp\left[-\frac{(z - x - \mu t)^2}{2\sigma^2 t}\right], \quad t > 0.$$

例 几何布朗运动 (7.4 节例 D) 经常被用来建立有价值东西的模型, 譬如一个完善的市场交易的股票价格. 这些价格都是非负的且通常表示出波动行为, 包含几何布朗运动的两个特征: 间歇的指数增长与从长远看呈指数衰减. 更重要地, 若 $t_0 < t_1 < \cdots < t_n$ 为时间点, 逐次比率

$$Y(t_1)/Y(t_0), \cdots, Y(t_n)/Y(t_{n-1}),$$

是独立随机变量, 因此, 粗略而言不重叠时间间隔的比率变化是相互独立的. 从粗略形式上讲, 这正是几何布朗运动能够在一个完善的市场中成为适当模型的原因所在. 如果未来价格跟当前价格的比率 $Y(t+s)/Y(t)$ 被预测是有利的, 很多买家将进入市场, 他们的要求倾向于提高当前价格 $Y(t)$. 类似地, 若 $Y(t+s)/Y(t)$ 被预测是不利的, 很多卖家将出现且倾向于降低当前价格. 当比率无法被预测是有利还是不利的时候, 也就是当不重叠时间间隔的价格比率相互独立的时候, 获得平衡.

我们将给出一个例子, 几何布朗运动被用来估计股票市场一种永久认股权证的价值. 一种认股权证是在一个特定时间段的任意时间以规定价格在一给定的股票市场购买固定数量的股票的选择权. 持有者的利润是市场价格高于选择价格的超出部分. 假定持有者可以以规定价格购买而以市场价格卖出以实现潜在利润.

我们只考虑没有期满日期的永久认股权证. 对于这种权证, 一个合理的策略是当股票价格首次达到某个特定的水平时 (记为 a) 就进行行权. 选择适当的单位, 我们可以假定权证的规定价格为 1, 从而, 当市场价为 a 时, 潜在利润为 $a-1$. 当然, 我们只考虑 $a>1$ 的情况, 因为人们不会在当前价格比 1 小的时候还以规定价格 1 购买股票.

363

当某人拥有这种权证时, 他就拥有前面所述的股票的直接所有权, 它以单位时间比率为 $\alpha = \mu + \frac{1}{2}\sigma^2$ 的速度增加, 因为

$$E[Y(t)|Y(0)=y] = y\exp\left\{t\left(\mu + \frac{1}{2}\sigma^2\right)\right\}.$$

人们需要有更高的回报率, $\theta > \alpha$, 或等价地, 以单位时间 $-\theta$ 的速率贴现潜在利润 $(a-1)$.

设 $T(a)$ 为股价第一次达到水平 a 的时间, 则权证持有者潜在利润的贴现为

$$\mathrm{e}^{-\theta T(a)}[Y(T(a))-1] = \mathrm{e}^{-\theta T(a)}(a-1).$$

我们想计算贴现利润的期望值, 从而选择 a 使利润的期望值达到最大. 根据布朗运动, $T(a)$ 是 $X(t) = \ln Y(t)$ 第一次达到水平 $\ln a$ 的时间, 在定理 5.3, 我们计算了 $T(a)$ 的概率密度函数, 以及在 (5.5) 式中计算了 $T(a)$ 的拉普拉斯变换. 利用 (5.5) 式, $z = \ln a$, $x = \ln y$, 我们有 (见图 7-4)

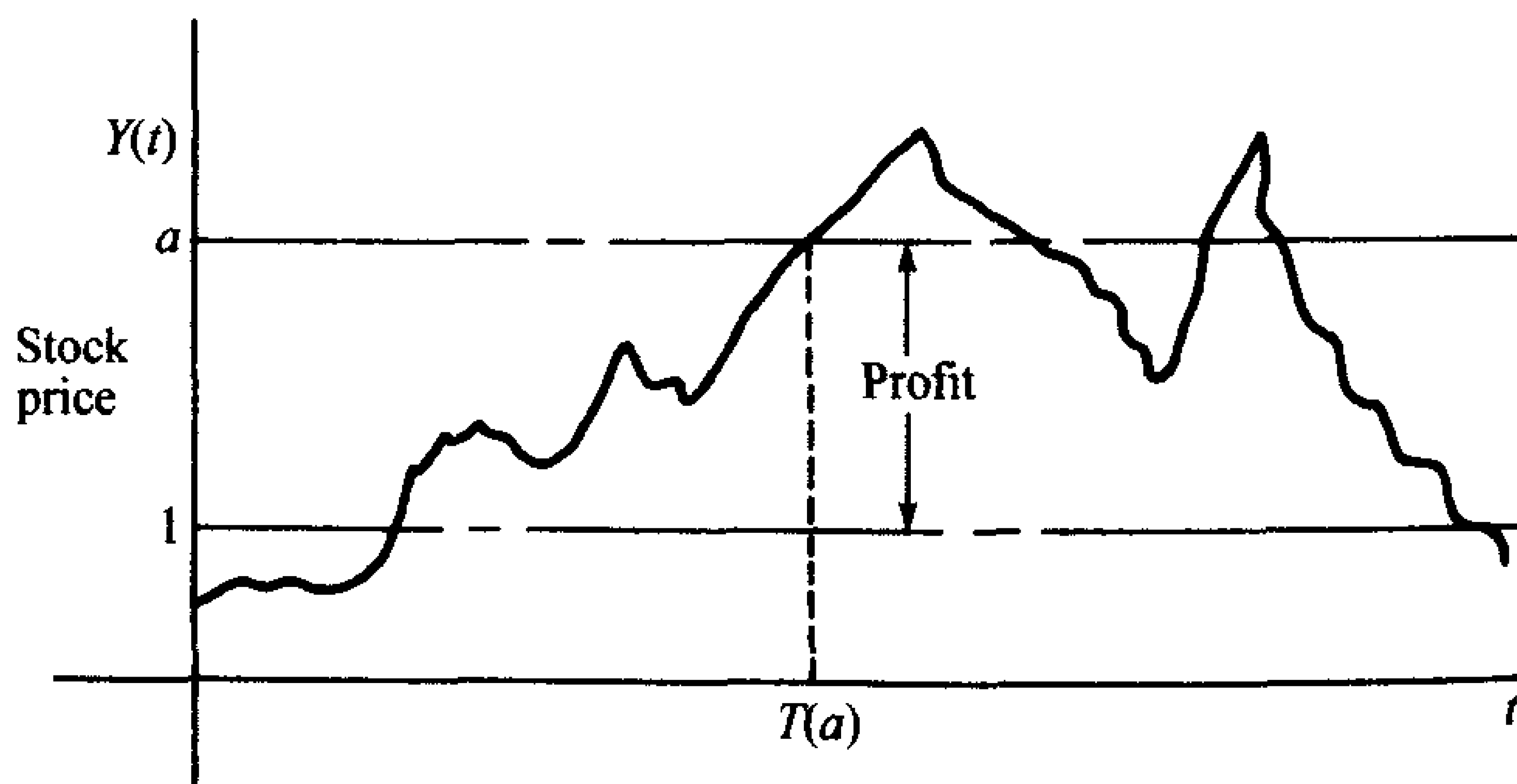


图 7-4

$$E[\mathrm{e}^{-\theta T(a)}|Y(0)=y]=\left(\frac{y}{a}\right)^{\rho},$$

364

其中

$$\rho=\sqrt{\frac{\mu^2}{\sigma^4}+\frac{2\theta}{\sigma^2}}-\frac{\mu}{\sigma^2}.$$

设 $g(y,a)$ 为期望的贴现利润, 我们有

$$\begin{aligned}g(y,a)&=(a-1)E[\mathrm{e}^{-\theta T(a)}|Y(0)=y]\\&=(a-1)\left(\frac{y}{a}\right)^{\rho}.\end{aligned}$$

我们将上式关于 a 微分并令其等于 0, 找到利润最大水平 $a=a^*$:

$$\frac{\partial g}{\partial a}=0=-\rho(a^*-1)\left(\frac{y}{a^*}\right)^{\rho+1}\frac{1}{y}+\left(\frac{y}{a^*}\right)^{\rho},$$

则

$$a^*=\frac{\rho}{\rho-1}.$$

条件 $\theta>\mu+\frac{1}{2}\sigma^2$ 保证 $1<a^*<\infty$. 给定当前股票价格 y, 则权证的价值为

$$\begin{aligned}g(y,a^*)&=(a^*-1)(y/a^*)^{\rho}\\&=\frac{1}{\rho-1}\left[\frac{y(\rho-1)}{\rho}\right]^{\rho}.\end{aligned}$$

7.6 多维布朗运动

设 $\{X_1(t);t\geqslant 0\},\cdots,\{X_N(t);t\geqslant 0\}$ 是标准布朗运动, 彼此间在下面意义下是

统计独立的：对于时间点的任何有限集合

$$t_{11}, t_{12}, \cdots, t_{1,n_1},$$
$$t_{21}, t_{22}, \cdots, t_{2,n_2},$$
$$\vdots$$
$$t_{N,1}, t_{N,2}, \cdots, t_{N,n_N},$$

N 个向量

$$\boldsymbol{X}_1 = (X_1(t_{11}), \cdots, X_1(t_{1,n_1})),$$
$$\boldsymbol{X}_2 = (X_2(t_{21}), \cdots, X_2(t_{2,n_2})),$$
$$\vdots$$
$$\boldsymbol{X}_N = (X_N(t_{N1}), \cdots, X_N(t_{N,n_N})),$$

365

是相互独立的. 向量值过程

$$\boldsymbol{X}(t) = (X_1(t), \cdots, X_N(t)), \quad t \geqslant 0,$$

称为 N 维布朗运动. 在平面上和在空间中布朗运动的粒子的运动可分别由二维和三维布朗运动过程来描述.

考虑二维布朗运动 $\boldsymbol{X}(t) = (X_1(t), X_2(t))$. 并计算当第一个坐标首次达到给定水平 $z > 0$ 时第二个坐标的分布. 令 T_z 是满足 $X_1(t) = z$ 条件的首次时间, 现在我们来求 $Y(z) = X_2(T_z)$ 的分布.

图 7-5 描绘了二维布朗运动在平面上的一条轨道, 并绘出了 $Y(z) = X_2(T_z)$ 的数值.

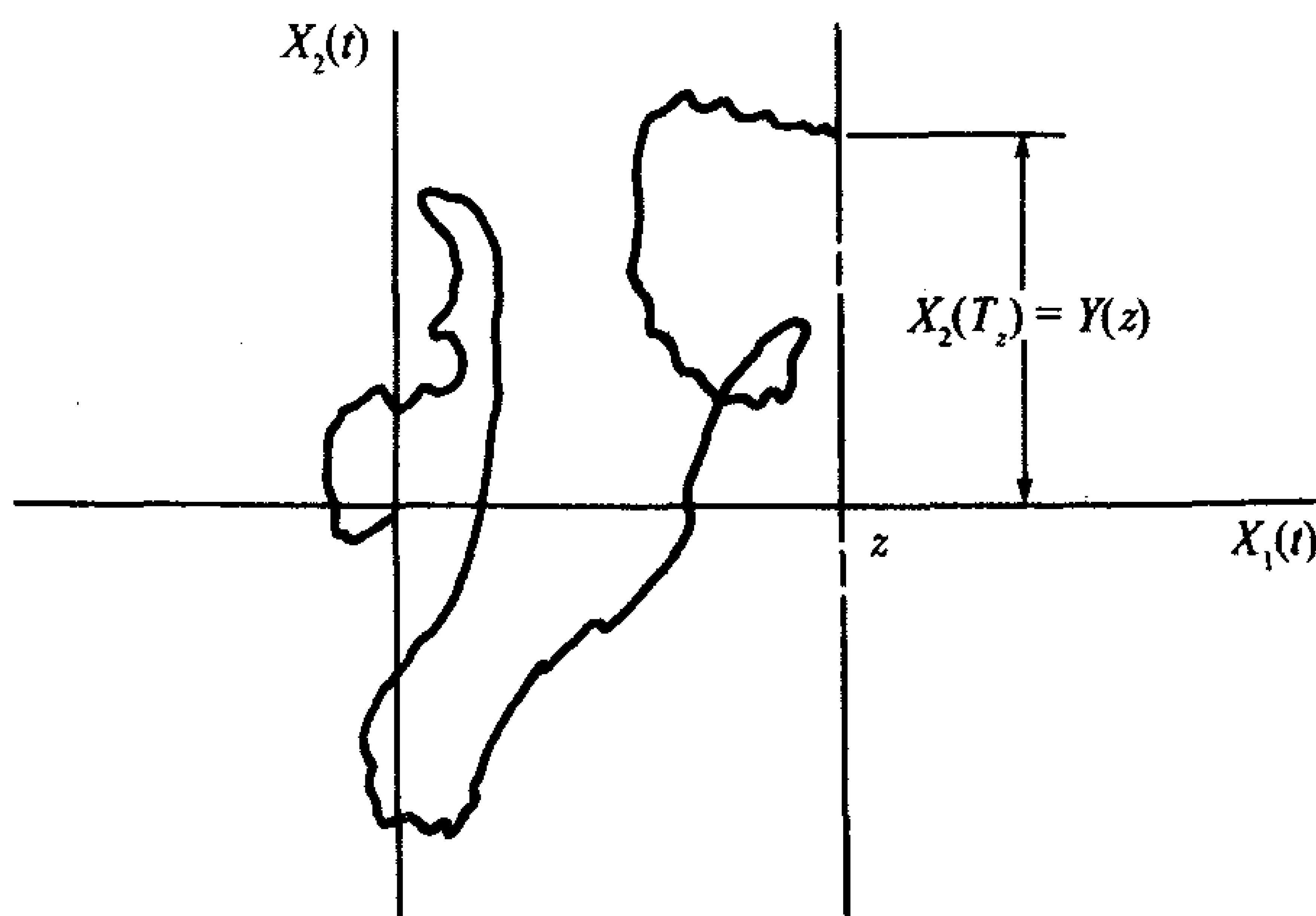


图 7-5 二维布朗运动

固定 $z>0$, 我们来计算特征函数 $\phi(u)=E[\exp\{\mathrm{i}uY(z)\}]$ 由于 T 是由过程 X_1 确定的, T 与过程 X_2 是独立的, 因此, 由全概率公式可知, 公式

$$\phi(u)=\int_0^\infty E[\exp\{\mathrm{i}uX_2(t)\}]\mathrm{d}F(t),$$

其中

$$F(t)=\Pr\{T_z\leqslant t\}$$

366

是 T_z 的累积分布函数, 可由等式 (3.3) 计算得到. 由于 $X_2(t)$ 是均值为 0 和方差为 t 的正态分布,

$$E[\exp\{\mathrm{i}uX_2(t)\}]=\exp\left(-\frac{1}{2}u^2t\right).$$

显然

$$\begin{aligned}\phi(u)&=\int_0^\infty \exp\left(-\frac{1}{2}u^2t\right)\mathrm{d}F(t)\\&=E\left[\exp\left(-\frac{1}{2}u^2T_z\right)\right]\\&=\exp(-|u|z),\qquad -\infty<u<\infty.\end{aligned}$$

上式是利用 T_z 的拉普拉斯变换 (见公式 (5.5)) 得到的, 其中 $\theta=\dfrac{1}{2}u^2,\mu=0$ 且 $\sigma^2=1$.

这是柯西概率密度函数

$$p(x)=\frac{1}{\pi z[1+(x/z)^2]},\quad -\infty<x<\infty.$$

的特征函数. 甚至我们可以得出更多结论. 初等问题 5 是证明随机过程 $\{T_z,z\geqslant 0\}$ 具有平稳独立增量的第一步. 由此推得

$$Y(z)=X_2(T_z)\tag{6.1}$$

也具有平稳独立增量. 一般来说, 若任给一个马尔可夫过程 $\{X(t),t\geqslant 0\}$ 和具有平稳独立增量和递增样本轨道的过程 $\{T_z;z\geqslant 0\}$, 其中 $T_0=0$, 我们就可引出新的过程

$$Y(z)=X(T_z),\quad z\geqslant 0.$$

从 X 形成 Y 的过程称为从属运算, 且过程 $\{T_z;z\geqslant 0\}$ 称为从属过程. 在上述条件下, $\{Y(z);z\geqslant 0\}$ 是马尔可夫过程. 此外, 如果 $\{X(t);t\geqslant 0\}$ 具有平稳独立增量, 则 $\{Y(z);z\geqslant 0\}$ 也同样具有平稳独立增量.

径向布朗运动

设 $\{\boldsymbol{X}(t); t \geqslant 0\}$ 是 N 维布朗运动, 过程

$$R(t) = [X_1(t)^2 + X_2(t)^2 + \cdots + X_N(t)^2]^{1/2}, \quad t \geqslant 0,$$

367 称为**径向布朗运动**, 或称具有参数 $\frac{1}{2}N - 1$ 的**贝塞尔过程**. 这是一个在状态空间 $[0, \infty)$ 有连续样本轨道的马尔可夫过程. 从 x 到 y 的转移概率密度函数是

$$p_t(x, y) = t^{-1} \exp\left\{-\frac{x^2 + y^2}{2t}\right\} (xy)^{1-(N/2)} I_{(N/2)-1}\left(\frac{xy}{t}\right) y^{N-1}, \quad t > 0, \quad x, y > 0, \tag{6.2}$$

其中 $I_v(z)$ 是修正贝塞尔函数:

$$I_v(z) = \sum_{k=0}^{\infty} \frac{(z/2)^{2k+v}}{k!\Gamma(k + v + 1)}. \tag{6.3}$$

对于 $N = 1$, 我们利用

$$I_{-1/2}(z) = \sqrt{\frac{2}{\pi z}} \cos\mathrm{h} z,$$

得到

$$p_t(x, y) = \sqrt{\frac{2}{\pi t}} \exp\left\{-\frac{x^2 + y^2}{2t}\right\} \cos\mathrm{h}\left(\frac{xy}{t}\right).$$

把这个公式与 (4.3) 比较, 可知当 $N = 1$ 时贝塞尔过程归结为反射布朗运动.

我们将简短地研究 $N = 2$ 的情况. 当 $N = 3$ 时, 关系式

$$I_{1/2}(z) = \sqrt{\frac{2}{\pi z}} \sin\mathrm{h} z$$

导出

$$p_t(x, y) = \sqrt{\frac{2}{\pi t}} \exp\left\{-\frac{x^2 + y^2}{2t}\right\} \frac{y}{x} \sin\mathrm{h}\left(\frac{xy}{t}\right).$$

令 $x \to 0$, 利用连续性我们得到对应于 $x = 0$ 的密度. 由此得到

$$p_t(0, y) = \sqrt{\frac{2}{\pi t^3}} y^2 \exp(-y^2/2t),$$

这是当 $N = 3$ 和 $R(0) = 0$ 时 $R(t)$ 的边缘密度.

现在我们来考虑 $N = 2$ 情况, 说明对应的贝塞尔过程是马尔可夫的、并计算其转移密度. 我们通过如下的极坐标变换:

$$R(t) = \sqrt{X_1(t)^2 + X_2(t)^2},$$

$$\Theta(t) = \arctan[X_2(t)/X_1(t)].$$

368

因为布朗运动是马尔可夫的,

$$\begin{aligned}&\Pr\{R(t_n+t)\leqslant b|\boldsymbol{X}(t_0)=\boldsymbol{x}_0,\cdots,\boldsymbol{X}(t_{n-1})=\boldsymbol{x}_{n-1},\boldsymbol{X}(t_n)=\boldsymbol{x}\}\\&\quad=\Pr\{R(t_n+t)\leqslant b|\boldsymbol{X}(t_n)=\boldsymbol{x}\}=\Pr\{R(t)\leqslant b|\boldsymbol{X}(0)=\boldsymbol{x}\},\end{aligned} \tag{6.4}$$

其中 $0<t_0<\cdots<t_n, t>0, \boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 是平面上任意的点, $\boldsymbol{x}=(x_1,x_2)$. 则

$$\begin{aligned}&\Pr\{R(t)\leqslant b|\boldsymbol{X}(0)=\boldsymbol{x}\}\\&=\iint_{y_1^2+y_2^2\leqslant b^2}\frac{1}{2\pi t}\exp\left\{-\frac{(y_1-x_1)^2+(y_2-x_2)^2}{2t}\right\}\mathrm{d}y_1\mathrm{d}y_2\\&=\int_0^b\left[\int_0^{2\pi}\frac{1}{2\pi t}\exp\left\{-\frac{(r\sin\theta-x_1)^2+(r\cos\theta-x_2)^2}{2t}\right\}\mathrm{d}\theta\right]r\mathrm{d}r,\end{aligned}$$

其中我们进行了变量替换 $y_1=r\sin\theta, y_2=r\cos\theta$. 并由微积分可知 $\mathrm{d}y_1\mathrm{d}y_2=r\mathrm{d}r\mathrm{d}\theta$. 由于

$$(r\sin\theta-x_1)^2+(r\cos\theta-x_2)^2=r^2-2r(x_1\sin\theta+x_2\cos\theta)+||\boldsymbol{x}||^2$$

其中 $||\boldsymbol{x}||^2=x_1^2+x_2^2$, 所以

$$\Pr\{R(t)\leqslant b|\boldsymbol{X}(0)=\boldsymbol{x}\}=\int_0^b\frac{r}{2\pi t}\exp\left\{\frac{r^2+||\boldsymbol{x}^2||}{2t}\right\}I(r,\boldsymbol{x})\mathrm{d}r$$

其中

$$I(r,\boldsymbol{x})=\int_0^{2\pi}\exp\left\{\frac{r}{t}(x_1\sin\theta+x_2\cos\theta)\right\}\mathrm{d}\theta.$$

设角 ϕ 满足

$$\sin\phi=x_1/||\boldsymbol{x}||,\quad \cos\phi=x_2/||\boldsymbol{x}||,$$

其中 $||\boldsymbol{x}||=\sqrt{x_1^2+x_2^2}$. 利用三角恒等式 $\sin\phi\sin\theta+\cos\phi\cos\theta=\cos(\phi+\theta)$,

$$\begin{aligned}I(r,\boldsymbol{x})&=\int_0^{2\pi}\exp\left\{\frac{r||\boldsymbol{x}||}{t}\cos(\phi+\theta)\right\}\mathrm{d}\theta\\&=\int_0^{2\pi}\exp\left\{\frac{r||\boldsymbol{x}||}{t}\cos\theta\right\}\mathrm{d}\theta,\end{aligned}$$

369

这是由于积分是在区间 $\theta\in[0,2\pi]$ 上, 在此区间 $\cos(\phi+\theta)$ 是周期的. 再者.

$$\begin{aligned}\int_0^{2\pi}\exp\{\alpha\cos\theta\}\mathrm{d}\theta&=\int_0^{2\pi}\sum_{k=0}^{\infty}\frac{(\alpha\cos\theta)^k}{k!}\mathrm{d}\theta\\&=\sum_{k=0}^{\infty}\frac{\alpha^k}{k!}\int_0^{2\pi}\cos^k\theta\mathrm{d}\theta.\end{aligned}$$

但

$$\int_0^{2\pi}\cos^k\theta\mathrm{d}\theta=\begin{cases}0, & \text{如果 } k=1,3,\cdots,\\ \dfrac{k!2\pi}{2^k[(k/2)!]^2}, & \text{如果 } k=0,2,\cdots,\end{cases}$$

这样, 利用修正贝塞尔函数, 有

$$\begin{aligned}\int_0^{2\pi}\exp\{\alpha\cos\theta\}\mathrm{d}\theta &= 2\pi\sum_{k=0,2,\cdots}\frac{\alpha^k}{k!}\left\{\frac{\alpha^k}{2^k[(k/2)!]^2}\right\}\\ &= 2\pi\sum_{j=0}^{\infty}\frac{\alpha^{2j}}{2^{2j}(j!)^2}\\ &= 2\pi I_0(\alpha).\end{aligned}$$

于是,

$$\begin{aligned}\Pr\{R(t)\leqslant b|\boldsymbol{X}(0)=\boldsymbol{x}\} &= \int_0^b\frac{r}{2\pi t}\exp\left\{\frac{r^2+||\boldsymbol{x}||^2}{2t}\right\}2\pi I_0\left\{\frac{r||\boldsymbol{x}||}{t}\right\}\mathrm{d}r\\ &= \int_0^b p_t(||\boldsymbol{x}||,r)\mathrm{d}r,\end{aligned}\tag{6.5}$$

其中 p_t 由等式 (6.2) 取 $N=2$ 确定. 这样, 我们证明了

$$\Pr\{R(t_n+t)\leqslant b|\boldsymbol{X}(t_0)=\boldsymbol{x}_0,\cdots,\boldsymbol{X}(t_n)=\boldsymbol{x}\}=\int_0^b p_t(||\boldsymbol{x}||,r)\mathrm{d}r.\tag{6.6}$$

令 $r_j=(x_{1j}^2+x_{2j}^2)^{1/2}, j=0,\cdots,n$, 其中 $\boldsymbol{x}_j=(x_{1j},x_{2j})$. 类似地, 令 $\theta_j=\arctan(x_{2j}/x_{1j})$. 条件 $\boldsymbol{X}(t_0)=\boldsymbol{x}_0,\cdots,\boldsymbol{X}(t_n)=\boldsymbol{x}_n$, 等价于条件 $R(t_0)=r_0$, $\Theta(t_0)=\theta_0,\cdots$, $R(t_n)=r_n,\Theta(t_n)=\theta_n$. 在式子 (6.6) 中利用这一点, 我们下面证明 $\{R(t),t\geqslant 0\}$ 的马尔可夫性并决定其转移密度函数. 基于全概率公式, 设 $p(\theta_0,\cdots,\theta_n)$ 是 $(\Theta(t_0),\cdots,\Theta(t_n))$ 的联合概率密度函数, 我们有

370

$$\begin{aligned}&\Pr\{R(t_n+t)\leqslant b|R(t_0)=r_0,\cdots,R(t_n)=r_n\}\\ &=\int_0^{2\pi}\cdots\int_0^{2\pi}\Pr\{R(t_n+t)\leqslant b|R(t_0)=r_0,\Theta(t_0)=\theta_0,\cdots,R(t_n)=r_n,\\ &\qquad\Theta(t_n)=\theta_n\}p(\theta_0,\cdots,\theta_n)\mathrm{d}\theta_0,\cdots,\mathrm{d}\theta_n\\ &\qquad=\int_0^{2\pi}\cdots\int_0^{2\pi}\left\{\int_0^b p_t(r_n,r)\mathrm{d}r\right\}p(\theta_0,\cdots,\theta_n)\mathrm{d}\theta_0,\cdots,\mathrm{d}\theta_n\\ &\qquad=\int_0^b p_t(r_n,r)\mathrm{d}r.\end{aligned}$$

这就验证了马尔可夫性并证实了式子 (6.2) 是二维贝塞尔过程的转移密度.

7.7 布朗运动的轨道

把布朗运动的样本路线或轨道看成随机选择的函数 (与作为随机变量集合的观点不同) 是值得我们注意的. 如果我们想得到一个处处不可微的连续函数, 我们必须付出巨大努力. 然而布朗运动的轨道肯定 (意指概率为 1) 是这样的函数! 这只不过是布朗运动轨道众多引人注目的特点之一.

A. 轨道的连续性

关于指标是实区间随机过程 $X(t)$ 的连续性, 有几种不同的类型. 其中的三种与第一章中介绍的随机序列三种极限概念相对应. 我们说 $\{X(t)\}$ 是

(a) 均方连续的, 如果对每个 t, 有

$$\lim_{s\to t} E[|X(s)-X(t)|^2]=0.$$

(b) 依概率连续的, 如果对每个 t 和正数 ε, 有

$$\lim_{s\to t}\Pr\{|X(s)-X(t)|>\varepsilon\}=0.$$

371

(c) 几乎必然连续的, 如果对于每个 t, 有

$$\Pr\{\lim_{s\to t} X(s)=X(t)\}=1.$$

前两个概念可由过程的有限维分布确定. 的确, 二阶矩有限的过程是均方连续的当且仅当均值函数 $m(t)=E[X(t)]$ 连续并且其协方差函数 $\Gamma(s,t)=E[\{X(s)-m(s)\}\{X(t)-m(t)\}]$ 在对角线 $t=s$ 处连续. 这可以从下面展开式中看出:

$$E[|X(s)-X(t)|^2]=\Gamma(s,s)-2\Gamma(s,t)+\Gamma(t,t)+[m(s)-m(t)]^2.$$

由切比雪夫不等式

$$\Pr\{|X(s)-X(t)|>\varepsilon\}\leqslant\frac{1}{\varepsilon^2}E[|X(s)-X(t)|^2],\quad \varepsilon>0,$$

可知每个均方连续过程均是依概率连续的.

虽然对于很多情况, 这些连续性概念是十分合理和有用的, 但对于所有情况它们并不都是适用的. 按照这三种意义, 泊松过程 $N(t)$ 都是连续的. 事实上, 首先注意到均值函数 $m(t)=\lambda t$ 和协方差 $\Gamma(s,t)=\lambda\min\{s,t\}$ 显然是连续的. 再者, 对固定的 t, 事件 $\lim_{s\to t}X(s)\neq X(t)$ 发生仅当过程在 t 时刻有跳跃. 既然这些跳跃时刻具有连续分布 (Γ 分布), 故这些事件的概率为 0. 这样, 对于每个 $t\geqslant 0$,

$$\Pr\{\lim_{s\to t}N(s)=N(t)\}=1,$$

因此, 泊松过程依照我们的定义, 在每个指定的 t 处也是几乎必然连续的.

可是, 我们从未见到一个泊松过程的轨道呈现出连续状态! 显然, 对于随机函数连续性我们尚需要更严格的准则. 我们称随机过程 $X(t)$ 具有连续的轨道, 如果 $X(t)$ 以概率 1 是 t 的连续函数. 由这个定义产生了若干技术上的困难, 在第 1 章我们通过指定所有有限维分布确定一个随机过程. 可是, 若我们仅利用这些有限维分布能否可以确定过程几乎必然具有连续轨道? 实际上, 若 $X(t)$ 几乎必然具有连续轨道, 对某个随机变量 τ, 我们定义

372

$$\widetilde{X}(t)=\begin{cases} X(t), & \text{如果 } t\neq\tau,\\ 0, & \text{如果 } t=\tau. \end{cases}$$

则 $\widetilde{X}(t)$ 并不出现连续轨道. 然而, 只要当 τ 具有连续分布且独立于 $\{X(t)\}, X(t)$ 和 $\widetilde{X}(t)$ 才具有相同的有限维分布. 于是, 我们只好减弱要求. 说一个由有限维分布确定的随机过程几乎必然具有连续轨道, 如果存在该过程的一个具体代表 $\{X(t)\}$, 这里, 它肯定 (即以概率 1) 是 t 的连续函数.

下面我们给出布朗运动 $\{X(t);0\leqslant t\leqslant 1\}$ 这样的一个代表. 注意, 我们仅考虑 [0, 1] 中的时间指标 t.

定义在 [0, 1] 上 Haar 函数如下:

$$H_1(t)=1,\qquad 0\leqslant t\leqslant 1,$$

$$H_2(t)=\begin{cases} 1, & 0\leqslant t<\dfrac{1}{2},\\ -1, & \dfrac{1}{2}\leqslant t\leqslant 1, \end{cases}$$

$$H_{2^n+1}(t)=\begin{cases} 2^{n/2}, & 0\leqslant t<2^{-(n+1)},\\ -2^{n/2}, & 2^{-(n+1)}\leqslant t\leqslant 2^{-n},\\ 0, & \text{其他}. \end{cases}$$

$$H_{2^n+j}(t)=H_{2^n+1}\left(t-\frac{j-1}{2^n}\right),\quad j=1,\cdots,2^n.$$

前 6 个函数如图 7-6 所示.

Schauder 函数 $S_k(t)$ 是 Haar 函数的积分, $S_k(t)=\displaystyle\int_0^t H_k(\tau)\mathrm{d}\tau$. 它们的图形像是一些小帐篷, 并且如果把它们画出, 可知

$$\max_{0\leqslant t\leqslant 1} S_{2^n+j}(t)=\left(\frac{1}{2}\right)^{(n+2)/2},\quad n=0,1,\cdots,0\leqslant j\leqslant 2^n-1,\tag{7.1}$$

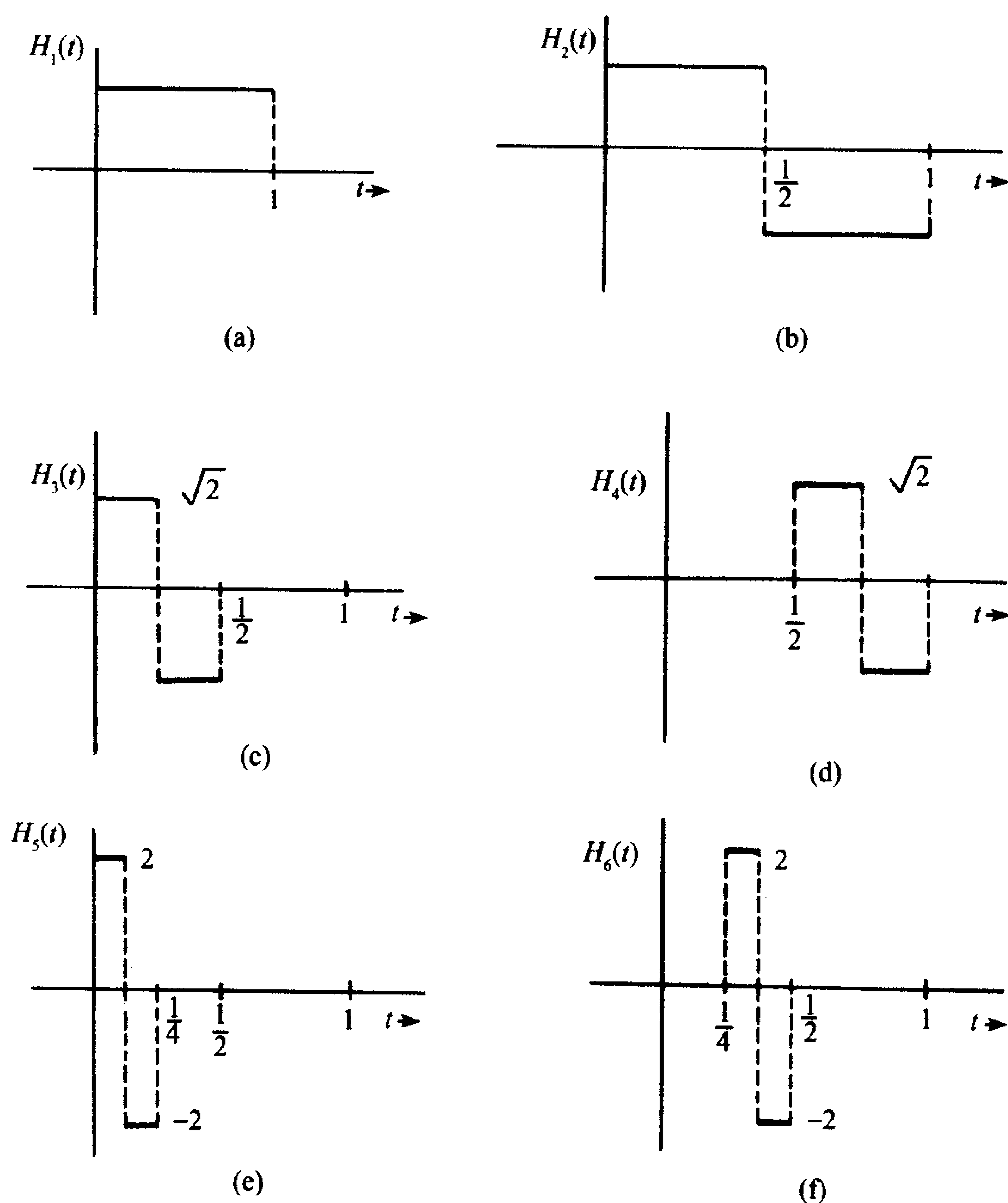


图 7-6

且

$$S_{2^n+j}(t)S_{2^n+k}(t)=0, \quad 1\leqslant k<j\leqslant 2^n. \tag{7.2}$$

现在令 $a(j), j=1,2,\cdots$ 为实数列, 置

$$b_n=\max\{|a(2^n+k)|; k=1,\cdots,2^n\}. \tag{7.3}$$

我们利用下面的事实. 如果

$$\sum_{n=0}^{\infty} b_n\left(\frac{1}{2}\right)^{n/2}<\infty, \tag{7.4}$$

则级数

$$x(t)=\sum_{k=1}^{\infty} a(k)S_k(t)$$

是一个一致收敛于 t 的连续函数. 为证明这个事实, 只要验证条件

$$\lim_{m,n\to\infty}\left|\sum_{k=m}^{m+n} a(k)\max_{0\leqslant t\leqslant 1} S_k(t)\right| = 0. \tag{7.5}$$

式 (7.1)~(7.3) 告诉我们

374

$$\left|\sum_{k=2^j+1}^{2^{j+1}} a(k)\max_{0\leqslant t\leqslant 1} S_k(t)\right| \leqslant b_j 2^{-(j+2)/2}, \tag{7.6}$$

所以通过把 (7.5) 中被加项依照指标 $k = 2^j, 2^j+1, \cdots, 2^{j+1}$ 分组, 我们可看到当 $m > 2^N$ 时, 在 (7.5) 中柯西和小于 $\sum_{j=N}^{\infty} b_j 2^{-(j+2)/2}$, 而后者在条件 (7.4) 之下收敛于 0. 那么, 当系数满足 (7.4) 时, $x(t)$ 是 t 的连续函数, 此即我们所要求的.

现在令 $A_1, A_2, A_3, \cdots$ 是具有零均值和单位方差的独立正态分布的随机变量序列. 令

$$B_n = \max\{|A_k| : 2^n < k \leqslant 2^{n+1}\}. \tag{7.7}$$

则当 $\sum_{n=0}^{\infty} B_n \left(\frac{1}{2}\right)^{n/2} < \infty$ 时,

$$X(t) = \sum_{k=1}^{\infty} A_k S_k(t) \tag{7.8}$$

是 t 的连续函数. 已知

$$\Pr\left\{\sum_{n=0}^{\infty} B_n \left(\frac{1}{2}\right)^{n/2} < \infty\right\} = 1,$$

所以, 式 (7.8) “必然” 定义一个连续函数. 为验证此断言, 我们首先对正态积分进行分部积分得

$$\begin{aligned}
\Pr\{|A_k| \geqslant x\} &= \frac{2}{\sqrt{2\pi}} \int_x^{\infty} \exp(-u^2/2)\mathrm{d}u \\
&= \frac{2}{\sqrt{2\pi}} \left\{\frac{\exp(-x^2/2)}{x} - \int_x^{\infty} \frac{\exp(-u^2/2)}{u^2}\mathrm{d}u\right\} \\
&\leqslant \frac{2}{\sqrt{2\pi}} \frac{\exp(-x^2/2)}{x}.
\end{aligned}$$

这样,

$$\sum_{n=2}^{\infty} \Pr\{|A_n| > 2\sqrt{\ln n}\} \leqslant k \sum_{n=2}^{\infty} \frac{1}{n^2\sqrt{\ln n}},$$

其中 k 是常数. 上面右边的和是收敛的, 由 Borel-Cantelli 引理 (第 1 章) 推得仅有有限多的值 $|A_n|$ 超过 $2\sqrt{\ln n}$. 这意味着仅有有限多的值 $B_j = \max\{|A_n| : 2^j < n \leqslant 2^{j+1}\}$ 超过 $2\sqrt{\ln 2}\sqrt{j}$. 由于 $\sum \sqrt{j}\left(\frac{1}{2}\right)^{j/2} < \infty$, 这就验证了 $\sum B_n \left(\frac{1}{2}\right)^{n/2}$ 是以概率 1 收敛的, 从而最终 $X(t) = \sum A_k S_k(t)$ 是连续的.

375

我们尚须证明 $X(t)$ 是布朗运动. 如果过程的每个有限维向量 $\{X(t_1), \cdots, X(t_k)\}$ 具有多元正态分布, 我们称其为**高斯过程**. 一个高斯过程是由它的均值和协方差函数所确定, 因为这些参数唯一地确定了所有有限维多元正态分布. 故为完成我们的证明, 我们只须证明 (1)$X(t)$ 是高斯的; (2)$E[X(t)] = 0$ 且 (3)$E[X(t)X(s)] = \min\{s, t\}$.

前面两个性质是容易验证的. 每个部分和 $X_n(t) = \sum_{k=0}^{n} A_k S_k(t)$ 是高斯的 (为什么?), 并且这个性质在极限情形仍保留. 确实, 容易验证 $X_n(t)$ 均方收敛于 $X(t)$, 这说明在式子 $E[X(t)] = \lim_{k\to\infty} \sum_{k=1}^{n} E[A_k] S_k(t) = 0$ 中, 极限号与期望的交换是合理的:

余下尚待确定的是 $X(t)$ 和 $X(s)$ 的协方差, 并看它与标准布朗运动的协方差 (即 $\min\{s, t\}$) 是否相等. 即, 我们希望验证等式

$$
\begin{aligned}
\min\{s, t\} &\overset{?}{=} E[X(s)X(t)] \\
&= \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} E[A_j A_k] S_j(s) S_k(t) \\
&= \sum_{k=1}^{\infty} S_k(s) S_k(t).
\end{aligned}
$$

这纯粹是古典分析中的问题. 但为了保持我们的论述自成体系, 并显示我们的方法和定理的威力, 下面应用鞅论证法. 在第 6 章问题 24 中, 我们曾断言

$$\sum_{k=1}^{n} a_k H_k(Z) = E[f(Z)|Y_1, \cdots, Y_n],$$

其中 $f(\tau)$ 是满足

$$\int_0^1 |f(\tau)| \mathrm{d}\tau < \infty, \quad a_k = \int_0^1 f(\tau) H_k(\tau) \mathrm{d}\tau \quad \text{和} \quad Y_k = H_k(Z)$$

的任意函数, Z 是在 $[0, 1]$ 上均匀分布的随机变量. 这样, 这些部分和构成 Doob 鞅, 因而以概率 1 收敛. 如果 $\int_0^1 |f(\tau)|^2 \mathrm{d}\tau < \infty$, 这个鞅还是平方可积的, 且当 $n \to \infty$ 时在均方收敛意义下有

$$\sum_{k=1}^{n} a_k H_k(Z) \to E[f(Z)|Y_1, Y_2, \cdots] = f(Z).$$

376

最后等式是由于 Z 被无限序列 $Y_1, Y_2, \cdots$ 所确定. 评估均差可见, 由于

$$\begin{aligned}
&\int_0^1 |f(\tau) - \sum_{k=1}^{n} a_k H_k(\tau)|^2 \mathrm{d}\tau \\
&= \int_0^1 \{f(\tau)\}^2 \mathrm{d}\tau - 2\sum_{k=1}^{n} a_k \int_0^1 f(\tau)H_k(\tau)\mathrm{d}\tau + \int_0^1 \left\{\sum_{k=1}^{n} a_k H_k(\tau)\right\}^2 \mathrm{d}\tau \\
&= \int_0^1 \{f(\tau)\}^2 \mathrm{d}\tau - \sum_{k=1}^{n} a_k^2.
\end{aligned}$$

且当 $n \to \infty$ 时, 上式收敛于 0, 我们推得 $\int_0^1 \{f(\tau)\}^2 \mathrm{d}\tau = \sum_{k=1}^{\infty} a_k^2$. 应用此公式于 $[f(\tau) + g(\tau)]$, 然后应用于 $f(\tau)$ 和 $g(\tau)$, 相减之后即得所谓的 Parseval 关系式:

$$\int_0^1 f(\tau)g(\tau)\mathrm{d}\tau = \sum_{k=1}^{\infty} a_k b_k, \tag{7.9}$$

其中 $b_k = \int_0^1 g(\tau)H_k(\tau)\mathrm{d}\tau$. 固定 $s < t$, 并令

$$f(\tau) = \begin{cases} 1, & 0 \leqslant \tau \leqslant s, \\ 0, & s < \tau \leqslant 1, \end{cases}$$

和

$$g(\tau) = \begin{cases} 1, & 0 \leqslant \tau \leqslant t, \\ 0, & t < \tau \leqslant 1. \end{cases}$$

则 $a_k = \int_0^1 f(\tau)H_k(\tau)\mathrm{d}\tau = S_k(s)$, 且 $b_k = S_k(t)$, 而当 $s < t$ 时有 $\int_0^1 f(\tau)g(\tau)\mathrm{d}\tau = s$. 代入 (7.9) 得 $S = \sum_{k=0}^{\infty} S_k(s)S_k(t)$. 对于 $t < s$ 可采用同样论证, 于是得到我们所希望的结果

$$\min\{s, t\} = \sum_{k=1}^{\infty} S_k(s)S_k(t).$$

这就完成了我们的证明, 即 $X(t) = \sum_{k=1}^{\infty} A_k S_k(t)$ 是布朗运动过程, 以概率 1 其轨道是 t 的连续函数.

对于指标集是 $0 \leqslant t < \infty$ 的布朗运动 $B(t)$, 首先对 $1 \leqslant t < \infty$ 令 $W(t) = tX(1/t)$, 置 $B(t) = W(1+t) - W(1), t \geqslant 0$. 我们让读者验证它确实是所希望的过程. (即它是高斯的且具有所希望的均值和方差.)

377

由上述构造法可推出布朗运动另一奇特性质. 将 (0, 1) 上均匀分布随机变量 U 的十进小数展开, $U = 0.Z_1Z_2Z_3\cdots$ 的数字按对角线方法排成如下的无限矩阵

$$\begin{array}{lllll} Z_1 & Z_3 & Z_6 & Z_{10} & \cdots \\ Z_2 & Z_5 & Z_9 & \cdots & \\ Z_4 & Z_8 & \cdots & & \\ Z_7 & \cdots & & & \end{array}$$

每一排给出的十进小数展开

$$\begin{aligned} &U_1 = 0.Z_1Z_3Z_6Z_{10}\cdots, \\ &U_2 = 0.Z_2Z_5Z_9Z_{14}\cdots, \\ &U_3 = 0.Z_4Z_8Z_{13}Z_{19}\cdots, \\ &\cdots \end{aligned}$$

是独立的并且在 (0, 1) 上均匀分布. 令 Φ^{-1} 是正态积分

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-t^2/2)\mathrm{d}t$$

的反函数, 并置 $A_k = \Phi^{-1}(U_k)$. 这些随机变量是相互独立的, 且由逆概率变换, 它们服从零均值和单位方差的正态分布. 在公式 (7.8) 中利用它们来构造一个布朗运动过程. 这样, 我们利用单个均匀分布随机变量 U 作为我们 "随机性" 的唯一来源, 构造了整个布朗运动! 从这个意义上, 布朗运动并不比随机地抽取一个数字的试验有更多的 "随机性"!

B. 平方变差

令 $X(t)$ 是标准布朗运动, 我们不再证明 $X(t)$ 是处处不可微分的, 虽然前面已指出此事为真. 然而, 我们要做什么才能给这个结论以支持? 对于每个固定 $t > 0$, 我们将建立

$$\lim_{n\to\infty} \sum_{k=1}^{2^n} \left[X\left(\frac{k}{2^n}t\right) - X\left(\frac{k-1}{2^n}t\right) \right]^2 = t. \tag{7.10}$$

这个极限既是在均方意义下, 也是在以概率 1 收敛 (或几乎必然) 的意义下进行的. [378]

对左边极限, 模仿初等微积分公式, 有

$$\int_0^t [\mathrm{d}X(\tau)]^2 = t = \int_0^t \mathrm{d}\tau.$$

它以微分的形式表达即是 $[\mathrm{d}X(t)]^2 = \mathrm{d}t$! 初看此式, 很自然感觉有点难以置信, 带着强烈期望, 小心地尝试进行分析验证, 以预防可能的错误. 在迅速进行检验之后, 并未发现什么错误. 事实上, 可赋予微分公式 $[\mathrm{d}X(t)]^2 = \mathrm{d}t$ 以精确的意义, 它不但是正确的, 而且是非常有用的. 然而, 有关这些内容将推延到第 15 章在介绍扩散过程时再讨论.

在证明之前, 我们先导出 (7.10) 的一个简单的推论:

$$\lim_{n\to\infty}\sum_{k=1}^{2^n}\left|X\left(\frac{k}{2^n}t\right)-X\left(\frac{k-1}{2^n}t\right)\right|=\infty. \tag{7.11}$$

换句话说, 布朗运动轨道的全变差是无限的 (以概率 1). 这已暗示了前述的轨道不可微的性质, 但这并不等于证明这个性质. 全变差的无限性可从下面不等式推得:

$$\sum_{k=1}^{2^n}\left|X\left(\frac{k}{2^n}t\right)-X\left(\frac{k-1}{2^n}t\right)\right|\geqslant\frac{\sum\limits_{k=1}^{2^n}\left[X\left(\frac{k}{2^n}t\right)-X\left(\frac{k-1}{2^n}t\right)\right]^2}{\max\limits_{j=1,\cdots,2^n}\left|X\left(\frac{j}{2^n}t\right)-X\left(\frac{j-1}{2^n}t\right)\right|}.$$

上面右边的分子趋于 t, 而分母趋向于 0, 因为布朗运动的轨道是连续的, 因此在有界区间内是一致连续的. 故左边部分趋于无穷, 此即说明 (7.11) 正确.

如果我们引入一些简明的记号, 平方变差公式的证明会变得方便得多. 令

$$\Delta_{nk}=X\left(\frac{k}{2^n}t\right)-X\left(\frac{k-1}{2^n}t\right),\quad k=1,\cdots,2^n,$$

且

$$W_{nk}=\Delta_{nk}^2-t/2^n,\qquad k=1,\cdots,2^n.$$

我们希望证明 $\sum\limits_{k=1}^{2^n}\Delta_{nk}^2\to t$, 或者相同地, $\sum\limits_{k=1}^{2^n}W_{nk}\to 0$. 对于每个 n, 在 $\{W_{nk}\}_{k=1}^{2^n}$ 中的随机变量是独立同分布的, 并且

379

$$E[W_{nk}]=E[\Delta_{nk}^2]-t/2^n=0, E[W_{nk}^2]=2t^2/4^n.$$

最后计算涉及正态分布 Δ_{nk} 的四阶矩. 事实上, 如果 Δ 是均值为 0, 方差为 σ^2 的正态分布, 则 $E[\Delta^{2m}]=1\times3\times\cdots\times(2m-1)\sigma^{2m}$, 此等式通过对正态特征函数进行微分容易给予归纳证明. 显然, 当 $j\neq k$ 时, 有 $E[W_{kn}W_{jn}]=0$, 故平方和式, 可得

$$E\left[\left\{\sum_{k=1}^{2^n}W_{kn}\right\}^2\right]=\sum_{k=1}^{2^n}E[W_{nk}^2]=2^{n+1}t^2/4^n=2t^2/2^n.$$

由于当 $n \to \infty$ 时, $2t^2/2^n \to 0$, 由此即可证明在均方收敛意义下所希望的平方变差公式. 为了得到以概率 1 收敛性, 令 $\varepsilon > 0$ 是任意给定的, 应用切比雪夫不等式知

$$\Pr\left\{\left|\sum_{k=1}^{2^n} W_{nk}\right| > \varepsilon\right\} \leqslant \frac{2t^2}{\varepsilon^2}\left(\frac{1}{2}\right)^n.$$

由于 $\sum\left(\frac{1}{2}\right)^n < \infty$, 依 Borel-Cantelli 引理知, 只有有限多个 n 的值, 能使 $\left|\sum_{k=1}^{2^n} W_{nk}\right| > \varepsilon$ 成立. 由于 $\varepsilon > 0$ 是任意的, 我们必定有 (以概率 1)

$$\lim_{n\to\infty}\sum_{k=1}^{2^n} W_{nk} = 0.$$

这等价于在几乎必然意义下平方变差关系式 (7.10).

尽管 $X(t)$ 是不可微分的, 在 9.8 节我们将研究诸如 $\int_0^t f(\tau)\mathrm{d}X(\tau)$ 的表达式的意义, 我们将定义积分为近似和的均方极限. 一个更加一般的随机积分将在第 15 章介绍.

C. 重对数律

著名的布朗运动重对数律最重要的形式指出

$$\limsup_{t\downarrow 0}\frac{X(t)}{\sqrt{2t\ln\ln(1/t)}} = 1 \tag{7.12}$$

为一个必然事件, 或者等价地, 是一个概率为 1 的事件. 它有多种变形和推广. 例如, 由于 $tX(1/t)$ 也是布朗运动, 我们有

$$\limsup_{t\downarrow 0}\frac{tX(1/t)}{\sqrt{2t\ln\ln(1/t)}} = 1,$$

380

或者, 令 $s = \frac{1}{t}$, 我们有

$$\limsup_{s\uparrow\infty}\frac{X(s)}{\sqrt{2s\ln\ln s}} = 1.$$

任给一个正数 ε, 一方面, 这意味着不管取多大的值 s, 必存在值 $t > s$, 使得

$$\frac{1}{\sqrt{t}}X(t) > (1-\varepsilon)\sqrt{2\ln\ln t},$$

而在另一方面, 对每条样本轨道, 只要选取充分大的 s, 即可以保证对所有 $t > s$ 有

$$\frac{1}{\sqrt{t}}X(t) < (1+\varepsilon)\sqrt{2\ln\ln t}.$$

在这种形式下, 重对数律对于下面本质上其实很简单的问题提供了一个令人注目的答案. 对任意固定的 t, $X(t)/\sqrt{t}$ 是均值为 0 和方差为 1 的正态分布, 所以

$$\Pr\left\{\frac{1}{\sqrt{t}}X(t) > k\sqrt{2\ln\ln t}\right\} = 1 - \Phi(k\sqrt{2\ln\ln t}),$$

其中 $\Phi(x)$ 是正态积分. 对于 $k=1$ 和 $t=10^{10}$, 这个概率是非常小, 大约是 0.006. 而另一方面, 推断下面事件: 是否存在某个 $t > 10^{10}$ 可使

$$\frac{1}{\sqrt{t}}X(t) > k\sqrt{2\ln\ln t}$$

的概率尽可能大? 如果 $k>1$, 重对数律作了否定的回答. 如果 $k<1$, 则是肯定的, 且其概率接近于 1.

我们将仅证明重对数律的一半, 即以概率 1 成立

$$\limsup_{t\downarrow 0} \frac{X(t)}{\sqrt{2t\ln\ln(1/t)}} \leqslant 1. \tag{7.13}$$

为此, 我们先对非负鞅

381

$$Z(s) = \exp\left\{\alpha X(s) - \frac{1}{2}\alpha^2 s\right\}, \alpha > 0$$

应用最大值不等式, 当 $\alpha > 0, \beta > 0$ 时, 推得

$$\begin{aligned}\Pr\left\{\max_{k=1,\cdots,2^n} \exp\left\{\alpha X\left(\frac{k}{2^n}t\right) - \frac{1}{2}\alpha^2\frac{k}{2^n}t\right\} > \mathrm{e}^{\alpha\beta}\right\} &\leqslant \mathrm{e}^{-\alpha\beta}E[Z(t)]\\ &= \mathrm{e}^{-\alpha\beta}E[Z(0)]\\ &= \mathrm{e}^{-\alpha\beta}.\end{aligned}$$

对每个 $n=1,2,\cdots$ 成立, 由于 $X(s)$ 的轨道是连续的, 令 $n\to\infty$, 得

$$\begin{aligned}\mathrm{e}^{-\alpha\beta} &\geqslant \Pr\left\{\sup_{0\leqslant s\leqslant t} \exp\left\{\alpha X(s) - \frac{1}{2}\alpha^2 s\right\} > \mathrm{e}^{\alpha\beta}\right\}\\ &= \Pr\left\{\sup_{0\leqslant s\leqslant t}\left\{X(s) - \frac{1}{2}\alpha s\right\} > \beta\right\}, \quad \alpha > 0.\end{aligned}$$

固定值 $\theta, 0<\theta<1$, 置

$$h(t) = \sqrt{2t\ln\ln(1/t)},$$

对 $\varepsilon > 0$, 选取

$$\alpha = \alpha_n = (1+2\varepsilon)\theta^{-n}h(\theta^n), \quad \beta = \beta_n = \frac{1}{2}h(\theta^n).$$

我们得到

$$\alpha\beta = (1+2\varepsilon)\ln\ln\theta^{-n} = (1+2\varepsilon)\ln nc, \quad 对于\ c = \ln(1/\theta) > 0.$$

这样, $e^{-\alpha\beta} = (nc)^{-(1+2\varepsilon)}$. 由于 $\sum_{n=1}^{\infty}(nc)^{-(1+2\varepsilon)} < \infty$, 应用 Borel-Cantelli 引理, 推得

$$\sup_{0\leqslant s\leqslant t}\left\{X(s) - \frac{1}{2}\alpha_n s\right\} \leqslant \beta_n, \tag{7.14}$$

上式对于除了有限多个值以外的所有 n 成立. 特别, 我们可以断言, 存在某个整数 N(由于它取决于所研究的特定轨道, 因而是随机的) 使得当 $n > N$ 时, (7.14) 成立. 如果 $t < \theta^N$, 那么, 当 $t \in (\theta^n, \theta^{n-1})$ 时, 我们有 $n > N$, 所以

$$\begin{aligned}\beta_n &\geqslant \sup_{0\leqslant s\leqslant t}\left\{X(s) - \frac{1}{2}\alpha_n s\right\}\\ &\geqslant X(t) - \frac{1}{2}\alpha_n t,\end{aligned}$$

且

$$\begin{aligned}X(t) &\leqslant \frac{1}{2}\alpha_n t + \beta_n\\ &\leqslant \frac{1}{2}\alpha_n\theta^{n-1} + \beta_n\\ &= \frac{1}{2}(1+2\varepsilon)\theta^{-1}h(\theta^n) + \frac{1}{2}h(\theta^n)\\ &= \frac{1}{2}\left\{\frac{1+2\varepsilon}{\theta} + 1\right\}h(\theta^n).\end{aligned}$$

382

由于 $h(t)$ 对于靠近 0 的 t 是一个递增函数, $h(\theta^n) \leqslant h(t)$, 即有

$$X(t) \leqslant \frac{1}{2}\left\{\frac{1+2\varepsilon}{\theta} + 1\right\}h(t).$$

这个不等式对于所有足够靠近原点的 t 均成立. 精确地说, 对所有 $t \leqslant \theta^N$ 是成立的. 于是, 以概率 1, 成立

$$\limsup_{t\downarrow 0}\frac{X(t)}{h(t)} \leqslant \frac{1}{2}\left\{\frac{1+2\varepsilon}{\theta} + 1\right\}.$$

由于我们对 θ 仅施加限制 $0 < \theta < 1$, 令 $\theta \to 1$, 可得

$$\limsup_{t\downarrow 0}\frac{X(t)}{h(t)} \leqslant 1 + \varepsilon$$

以概率 1 成立. 既然 ε 是任意正数, 故我们证明了 (7.13) 所叙述的重对数律的一半.

初等问题

在下面问题中, $X(t)$ 均为标准布朗运动.

1. 设 T_0 是不超过 t 的 $X(\tau)$ 最大零点, 试证明

$$\Pr\{T_0 < t_0\} = \frac{2}{\pi}\arcsin\sqrt{t_0/t}.$$

提示: 利用定理 3.1.

2. 设 T_1 是超过 t 的 $X(t)$ 最小零点, 试证明

(a) $\Pr\{T_1 < t_1\} = \dfrac{2}{\pi}\arccos\sqrt{t/t_1}$.

(b) $\Pr\{T_0 < t_0, T_1 > t_1\} = \dfrac{2}{\pi}\arcsin\sqrt{t_0/t_1}$.

3. 试验证 $E(X(t)X(s)|X(0)=0) = \min(t,s)$.

4. 试证明密度函数

$$p(t,x,y) = \frac{1}{\sqrt{2\pi t}}\exp[-(x-y)^2/2t]$$

满足热方程

383

$$\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial x^2}.$$

5. 设 $T(\lambda)$ 是 $X(0)=0$ 条件下首次到达 $\lambda>0$ 的时间. 试证明 $T(\lambda_1+\lambda_2)$ 的分布与 $T(\lambda_1)+T(\lambda_2)$ 的分布相同, 其中把 $T(\lambda_1)$ 和 $T(\lambda_2)$ 看为独立的随机变量, $\lambda_1,\lambda_2>0$.

提示: 验证 $\phi_{\lambda_1+\lambda_2}(\theta) = \phi_{\lambda_1}(\theta)\phi_{\lambda_2}(\theta)$, 其中 $\phi_\lambda(\theta)$ 是由等式 (5.3) 所给的 $T(\lambda)$ 的拉普拉斯变换.

6. 试确定

$$U(t) = \mathrm{e}^{-t}X(\mathrm{e}^{2t}), \quad t \geqslant 0,$$

和

$$V(t) = X(t) - tX(1), \quad 0 \leqslant t \leqslant 1$$

的协方差函数.

解答:

$$E[U(t)U(s)] = \exp\{-|t-s|\},$$

和

$$E[V(t)V(s)] = t(1-s), t \leqslant s.$$

7. 对于标准布朗运动 $X(t)$ 和常数 $\alpha>0,\beta>0$, 试证明 $\Pr\{X(t) \leqslant \alpha t+\beta$, 对所有 $t \geqslant 0|X(0)=w\} = 1-\mathrm{e}^{-2\alpha(\beta-w)}$, 对于 $w \leqslant \beta$.

提示: 应用推论 5.1 于漂移布朗运动 $W(t) = X(t)-\alpha t - w$.

8. 设 $W = \int_0^t X(s)\mathrm{d}s$, 试证明 $E[W]=0$ 和 $E[W^2] = t^3/3$.

提示：验证

$$\begin{aligned}E[W^2]&=E\left[\left\{\int_0^t X(s)\mathrm{d}s\right\}^2\right]\\&=E\left[\left\{\int_0^t X(u)\mathrm{d}u\right\}\left\{\int_0^t X(v)\mathrm{d}v\right\}\right]\\&=\int_0^t\int_0^t E[X(u)X(v)]\mathrm{d}u\mathrm{d}v\\&=2\int_0^t\left\{\int_0^v u\mathrm{d}u\right\}\mathrm{d}v.\end{aligned}$$

9. 试推导在已知 $X(t)=x$ 条件下 $W=\int_0^t X(s)\mathrm{d}s$ 的条件分布.

提示：W 和 $X(t)$ 具有联合正态分布.

解答：已知 $X(t)=x$ 条件下, W 是具有均值 $E[W|X(t)=x]=\frac{1}{2}tx$ 和方差 $E\left[\left(W-\frac{1}{2}tx\right)^2|X(t)=x\right]=t^3/12$ 的正态分布.

384

10. 设 T 是布朗运动过程首穿直线 $l(t)=\alpha+\beta t(\alpha>0,\beta>0)$ 的时间, 试求 T 的拉普拉斯变换.

提示：利用 (5.1) 的第二个鞅, 得

$$E\left[\exp\left\{\lambda(\alpha+\beta T)-\frac{1}{2}\lambda^2 T\right\}\right]=1$$

然后进行变量替换 $\lambda\beta-\frac{1}{2}\lambda^2=z$.

11. 设 $Y(t)=\mathrm{e}^{X(t)}$ 是几何布朗运动, 试求扩散系数

$$\lim_{h\downarrow 0}\frac{E[Y(t+h)-Y(t)|Y(t)=y]}{h}=b(y),\quad 0<y<\infty,$$

和

$$\lim_{h\downarrow 0}\frac{E[\{Y(t+h)-Y(t)\}^2|Y(t)=y]}{h}=a(y),0<y<\infty.$$

12. 利用关系式 (5.5) 求下面积分的值

$$\int_0^\infty\frac{1}{\sqrt{t}}\exp\left\{-\left(at+\frac{b}{t}\right)\right\}\mathrm{d}t,\qquad a,b>0;$$

$$\int_0^\infty\frac{1}{t^{3/2}}\exp\left\{-\left(at+\frac{b}{t}\right)\right\}\mathrm{d}t.$$

13. 试证明：$\Pr\{M(t)>\delta|X(t)=M(t)\}=\exp(-\delta^2/2t)$, 其中 $M(t)=\max\limits_{0\leqslant u\leqslant t}X(u)$.

提示：令 $Y(t)=M(t)-X(t)$. 在 $Y(t)=0$ 条件下, 计算 $M(t)$ 的条件分布.

14. 试证明恒等式

(i) $E\left[\exp\left\{\lambda\int_0^t X(s)\mathrm{d}s\right\}\right]=\exp(\lambda^2t^3/6),\quad -\infty<\lambda<\infty,$

(ii) $E\left[\exp\left\{\lambda\int_0^t sX(s)\mathrm{d}s\right\}\right] = \exp(\lambda^2 t^5/15), \quad -\infty < \lambda < \infty.$

15. 设 $R(t) = [X_1(t)^2 + \cdots + X_m(t)^2]^{1/2}$ 是径向布朗运动或 m 维贝塞尔过程. (a) 试证 $R(t)^2 - mt$ 是鞅. (b) 利用鞅可选抽样定理证明 $E[T] = \gamma^2/m$, 其中 $T = \inf\{t \geqslant 0; R(t) \geqslant \gamma\}$ 是 m 维布朗运动 $[X_1(t), \cdots, X_m(t)]$ 首次到达距离原点为 γ 的时间.

385

问 题

以下我们使用记号

$$M(t) = \max_{0 \leqslant u \leqslant t} X(u),$$

和

$$Y(t) = M(t) - X(t),$$

其中 $X(t)$ 是标准布朗运动.

1. 试证明 $Y(t) = M(t) - X(t)$ 是连续时间马尔可夫过程.

提示: 注意, 对于 $t' < t$ 有

$$Y(t) = \max\{\max_{t' \leqslant u \leqslant t}(X(u) - X(t')), Y(t')\} - (X(t) - X(t')).$$

2. 试证明随机过程 $Y(t) = M(t) - X(t)$ 和随机过程 $|X(t)|$ 是等价的. (两个过程等价是指它们的有限维分布相同.)

提示: 由于 $|X(t)|$ 和 $Y(t)$ 皆为马尔可夫过程, 因此只要证明密度函数

$$\Pr\{Y(t) < y | Y(t_0) = y_0, t_0 < t\} \text{ 和 } \Pr\{|X(t)| < y | |X(t_0)| = y_0, t_0 < t\}$$

是恒等的.

可利用问题 1 的 $Y(t)$ 表示式计算上面左式.

3. 试证明 $Y(t)$ 在区间 (t_0, t_1) 至少有一个零点的概率是 $(2/\pi)\arccos\sqrt{t_0/t_1}$.

4. 设 $T_1^*(T_0^*)$ 是 $Y(\tau) = M(\tau) - X(\tau)$ 超过 (不超过)t 的最小 (最大) 零点. 试证明 T_0^* 和 T_1^* 分别具有与 T_0 和 T_1 相同的分布. T_0 和 T_1 定义见初等问题 1 和 2.

5. 对于 $ab > 0$, 试证

$$\Pr\{X(t) \text{ 在 } (0,t) \text{ 不为 } 0 | X(0) = a, X(t) = b\} = 1 - \mathrm{e}^{-2ab/t}.$$

提示: 利用 (3.9) 的函数 $A_t(x, y)$.

6. 对于 $\alpha, \beta > 0$, 试证

$$\Pr\{X(u) < \alpha u + \beta, 0 \leqslant u \leqslant 1 | X(0) = X(1) = 0\} = 1 - \mathrm{e}^{-2\beta(\beta+\alpha)}.$$

提示: 利用定理 2.1 证明恒等式

$$\begin{aligned}&\Pr\{X(u) < \alpha u + \beta, 0 \leqslant u \leqslant 1 | X(0) = X(1) = 0\}\\&= \Pr\{X(u) < 0, 0 \leqslant u \leqslant 1 | X(0) = -\beta, X(1) = -\beta - \alpha\},\end{aligned}$$

然后参考虑问题 5.

386

7. 已知 $X(t)$ 在区间 (t_0,t_1) 上不为零, 试求其在区间 (t_0,t_2) 上不为零的条件概率, 其中 $0<t_0\leqslant t_1\leqslant t_2$.

答案:

$$\frac{\arcsin\sqrt{t_0/t_2}}{\arcsin\sqrt{t_0/t_1}}.$$

8. 试证明在已知 $X(t)$ 于区间 $(0,t_1)$ 不为零条件下, 在区间 $(0,t_2)$ 不为零的条件概率是 $\sqrt{t_1/t_2}$, 其中 $0<t_1<t_2$.

提示: 计算

$$\Pr\{X(t)\neq 0, 0<t_0\leqslant t\leqslant t_2|X(t)\neq 0, 0<t_0\leqslant t\leqslant t_1\},$$

然后令 $t_0\to 0$.

9. 试证明: 已知 $X(t)$ 于区间 (t_0,t_1) 上在点 t_0 或 t_1 取到一极值 [注: $X(t)$ 有两个极值], 事件 $|X(t_1)-X(t_0)|>\xi$ 的条件概率是 $\exp(-\xi^2/2(t_1-t_0)), t_0>0$.

提示: 证明下面命题

(i) 问题中所述事件可在下面四种情况之一下发生: (A) $X(t_0)$ 为最小, (B) $X(t_0)$ 为最大, (C) $X(t_1)$ 为最小, (D) $X(t_1)$ 为最大.

(ii) 已知 (A), (B), (C), (D) 中任一个情况下, 其他任一情况发生的条件概率为 0.

(iii) $\Pr\{|X(t_1)-X(t_0)>\xi|$ (A), (B), (C), 或 (D) 发生$\}$

$$\begin{aligned}&=\sum_{\alpha=(\mathrm{A}),(\mathrm{B}),(\mathrm{C}),(\mathrm{D})}\Pr\{|X(t_1)-X(t_0)|>\xi|\alpha\}\\&\quad\times\Pr\{\alpha|(\mathrm{A}),(\mathrm{B}),(\mathrm{C}),\text{ 或 }(\mathrm{D})\text{ 发生}\}\\&=\exp[-\xi^2|2(t_1-t_0)].\end{aligned}$$

(利用初等问题 13 和反射原理)

10. 试证明

$$\begin{aligned}&\Pr\{X(\tau)\neq 0, 0<t<\tau<u<1|X(0)=X(1)=0\}\\&=\frac{2}{\pi}\arccos\sqrt{\frac{u-t}{u(1-t)}}.\end{aligned}$$

提示: 计算

$$\begin{aligned}&2\int_{\alpha=0}^{\infty}\int_{\tau=u}^{1}\Pr\{X(t)=\alpha, T(\alpha)=\tau-t, X(1-\tau)=0|X(0)=0\}\\&\times[\Pr\{X(1)=0|X(0)=0\}]^{-1}\mathrm{d}\alpha\mathrm{d}\tau,\end{aligned}$$

387

其中 $T(\alpha)$ 表示布朗粒子从 $X(0)=\alpha$ 出发首达 0 的时间 [见 (3.7)], 和

$$\frac{\mathrm{d}}{\mathrm{d}u}\left[\frac{2}{\pi}\arccos\sqrt{\frac{u-t}{u(1-t)}}\right]$$

$$= -\frac{2}{\pi}\frac{1}{\sqrt{1-(u-t)/u(1-t)}}\left(\frac{1}{\sqrt{1-t}}\right)\frac{t}{2\sqrt{1-(t/u)u^2}}$$

$$= -\frac{\sqrt{t}}{\pi}\frac{1}{(\sqrt{1-u})(\sqrt{u-t})u}.$$

11. 证明恒等式

$$E\left[\exp\left\{\lambda\int_0^t f(s)X(s)\mathrm{d}s\right\}\right]$$

$$= \exp\left\{\lambda^2\int_0^t f(v)\left[\int_0^v uf(u)\mathrm{d}u\right]\mathrm{d}v\right\}, \quad -\infty < \lambda < \infty,$$

其中 $f(s)$ 为任意连续函数, $0 \leqslant s < \infty$.

12. 证明 $\Pr\{X(1) \leqslant x | X(u) \geqslant 0, 0 \leqslant u \leqslant 1\} = 1 - \exp(-x^2/2)$.

提示: $X'(t) = X(1) - X(1-t)$ 也是布朗运动. 根据 $X'(t)$ 所求概率可表示为

$$\Pr\{X'(1) \leqslant x | M'(1) = X'(1)\},$$

其中 $M'(t) = \max\limits_{0\leqslant x\leqslant t} X'(x)$, 然后参考初等问题 13.

13. 对于 $a > 0, b < a$, 证明

$$\Pr\left\{\sup_{t\geqslant 0}\frac{b+X(t)}{1+t} \geqslant a\right\} = \mathrm{e}^{-2a(a-b)}.$$

然后证明左边部分, 因此也是右边部分, 等于

$$\Pr\{\sup_{0\leqslant u\leqslant 1} X(u) \geqslant a | X(1) = b\}.$$

14. 试证明布朗运动的科尔莫戈罗夫不等式:

$$\Pr\{\sup_{0\leqslant u\leqslant t}|X(u)| > \varepsilon\} \leqslant t/\varepsilon^2, \quad \varepsilon > 0.$$

15. (续上) 利用科尔莫戈罗夫不等式证明

$$\lim_{t\to\infty}\frac{1}{t}X(t) = 0.$$

388

提示: 置 $\varepsilon = 2^{2n/3}$ 和 $t = 2^n$, 并利用 Borel-Cantelli 引理.

16. 对于 $n = 1, 2, \cdots$ 和 $k = 1, \cdots, 2^n$, 置

$$\Delta_{nk} = X\left(\frac{k}{2^n}\right) - X\left(\frac{k-1}{2^n}\right)$$

其中 $X(t)$ 是标准布朗运动. 试证明 $E[S_{n+1}|S_n] = \frac{1}{2}(s_n + 1)$, 其中 $S_n = \sum\limits_{k=1}^{2^n}\Delta_{nk}^2$.

提示: 利用定理 2.1 证明

$$E[\Delta_{n+1,2k-1}^2 + \Delta_{n+1,2k}^2 | X(j/2^n), j = 1, \cdots, 2^n] = \frac{1}{2}(\Delta_{nk}^2 + 1).$$

然后两边求和.

17. 利用问题 16 的记号, 证明 $E[S_n|S_{n+1}] = S_{n+1}$.

提示: 利用对称性证明

$$\begin{aligned}
&E[\Delta_{nk}^2|\Delta_{n+1,2k-1}^2, \Delta_{n+1,2k}^2] \\
&= E[(\Delta_{n+1,2k-1} + \Delta_{n+1,2k})^2|\Delta_{n+1,2k-1}^2, \Delta_{n+1,2k}^2] \\
&= \frac{1}{2}(\Delta_{n+1,2k-1} + \Delta_{n+1,2k})^2 + \frac{1}{2}(\Delta_{n+1,2k-1} - \Delta_{n+1,2k})^2 \\
&= \Delta_{n+1,2k-1}^2 + \Delta_{n+1,2k}^2.
\end{aligned}$$

18. 设 $X(t)$ 是标准布朗运动, 对于 $\varepsilon > 0$ 和 $T > 1$, 令 $g_{\varepsilon,T}(x)$ 表示在已知 $X(t) \geqslant -\varepsilon$(对所有 $t \leqslant T$) 之下 $X(1)$ 的条件概率密度. 试证明

$$\lim_{\substack{\varepsilon \downarrow 0 \\ T \to \infty}} g_{\varepsilon,T}(x) = \sqrt{\frac{2}{\pi}} x^2 \exp(-x^2/2).$$

注: 这是三维贝塞尔过程中 $R(1)$ 的分布.

19. 设 $f_\theta(x,t) = \exp\left\{\theta x - \frac{1}{2}\theta^2 t\right\}$, 于是 $\{f_\theta(X(t),t\}$ 对任意实参数 θ 是鞅. 利用鞅 $f_\theta(X(t),t) + f_{-\theta}(X(t),t)$, 其中 $\theta = \sqrt{2\lambda}$, 试证明

$$E[\mathrm{e}^{-\lambda T}] = \frac{1}{\cosh(\sqrt{2\lambda} a)},$$

此处 $T = \min\{t : X(t) = +a$ 或 $X(t) = -a\}$.

20. 置

$$p(x,t) = \frac{1}{\sqrt{t}} \exp(x^2/2t), t > 0.$$

试证明 $p(X(t), a+t)$ 对于任意 $a > 0$ 是鞅.

提示: 验证

$$p(x,t) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} f_\theta(x,t) \mathrm{d}\theta,$$

其中 $f_\theta(x,t) = \exp\left\{\theta x - \frac{1}{2}\theta^2 t\right\}$. 也可参考 (5.2).

389

21. 利用问题 20 中的鞅和鞅的最大值不等式证明

$$\Pr\{|X(t)| \geqslant \sqrt{2(a+t)\ln\sqrt{a+t}}, \quad 对于某个\ t \geqslant 0\} \leqslant \frac{1}{\sqrt{a}}.$$

22. 固定 $a < 0 < b$, 设 $T(a)$[分别地, $T(b)$] 是过程首达 a(分别地, 首达 b) 的时间. 令 $I_a = 1$ 若 $T(a) < T(b)$; 其他为 0. 类似地定义 I_b 为到达 a 之前到达 b 的事件的示性函数. 试证:

$$\exp(-\sqrt{2\lambda} b) = E\{I_b \exp[-\lambda T(b)]\} + E\{I_a \exp[-\lambda T(a)]\} \exp[-\sqrt{2\lambda}(b-a)],$$

和

$$\exp(\sqrt{2\lambda} a) = E\{I_a \exp[-\lambda T(a)]\} + E\{I_b \exp[-\lambda T(b)]\} \exp[-\sqrt{2\lambda}(b-a)].$$

提示：第一个方程是将到达 b 的轨道按照是否首先到达 a 分成两类. 然后利用

$$\exp(-\sqrt{2\lambda}b) = E[\exp[-\lambda T(b)]|X(0)=0]$$

和

$$\exp[-\sqrt{2\lambda}(b-a)] = E[\exp[-\lambda T(b)]|X(0)=a].$$

23. (续上) 从问题 22 所导出的方程组中求出 $E\{I_a \exp[-\lambda T(a)]\}$ 和 $E\{I_b \exp[-\lambda T(b)]\}$. 令 $T = \min\{T(a), T(b)\}$ 是首达 a 或 b 的时间. 求

$$E[\exp(-\lambda T)] = E\{I_a \exp[-\lambda T(a)]\} + E\{I_b \exp[-\lambda T(b)]\}.$$

24. 设 $W(t)$ 是带有正漂移参数 $\mu > 0$ 和方差 σ^2 的布朗运动. 令 $M(t) = \max\limits_{0\leqslant u\leqslant t} W(u)$ 和 $Y(t) = M(t) - W(t)$. 固定 $a > 0$ 和 $y > 0$, 并令

$$T(a) = \min\{t : M(t) = a\}, \quad S(y) = \min\{t : Y(t) = y\}.$$

试证明

$$\Pr\{T(a) < S(y)\} = \exp\left\{\frac{-2\mu a}{\sigma^2[\exp(2\mu y/\sigma^2) - 1]}\right\}.$$

提示：令 $f(a) = \Pr\{T(a) < S(y)\}$. 首先论证 $f(a_1 + a_2) = f(a_1)f(a_2)$, 因而对某个常数 k 有 $f(a) = \mathrm{e}^{-ka}$.

若 $\lambda = -2\mu/\sigma^2, \mathrm{e}^{\lambda X(t)} = \mathrm{e}^{\lambda[M(t)-Y(t)]}$ 是鞅. 置

$$T = \min\{T(a), s(y)\}$$

并应用可选停止定理. 注意 $M(T) = a, Y(T) = 0$, 若 $T(a) < S(y)$; $Y(T) = y$, 若 $T(a) > S(y)$. 根据 $T(a) < S(y)$ 及 $T(a) > s(y)$ 两种情况来处理 $1 = E[\mathrm{e}^{\lambda[M(T)-Y(T)]}]$. 令 $a \to 0$ 确定未知常数 k.

25. 设 $\{X(t), t \geqslant 0\}$ 是布朗运动, 关于 θ 微分鞅

390

$$L_\theta(t) = \exp\{\theta X(t) - (1/2)\theta^2 t\},$$

证明对每个 n, $H_n(X(t), t)$ 是鞅, 其中

$$H_0(x,t) \equiv 1,$$

$$H_1(x,t) = x,$$

和

$$H_n(x,t) = xH_{n-1}(x,t) - (n-1)tH_{n-2}(x,t).$$

另一种方法可利用公式 (5.2) 完成.

26. 考虑定义在实直线上满足下述条件的连续可积函数 f:

$$\int_{-\infty}^{\infty} f(\delta)\mathrm{d}\delta = a > 0.$$

构造过程

$$Y(t) = \frac{1}{\sqrt{t}} \int_0^t f(X(u))\mathrm{d}u.$$

试证明

$$\lim_{t\to\infty} E[Y(t)] \quad 和 \quad \lim_{t\to\infty} E[Y^2(t)]$$

存在并求出它们的值.

27. (续上) 试证明

$$\lim_{t\to\infty} E[\{Y(t)\}^k] = \mu_k a^k$$

其中 μ_k 是单边正态分布的 k 阶矩. (单边正态分布是 $|Z|$ 的分布, 其中 Z 服从标准正态分布.)

附　记

有关应用布朗运动于统计力学和数学分析的知识学习, 我们推荐 kac 所写的专著 [2]

关于扩散过程杰出的专题论文是 Ito 和 Mckean[3], 他们完善并深入地推广了 Lévy 的工作.

参考书目

[1] P. Lévy, *Processus Stochastiques et Mouvement Brownien*. Gauthier-Villars, Paris, 1948.

[2] M. Kac, *Probability and Related Topics in Physical Sciences*. Wiley, New York, 1959.

[3] K. Ito and H. P. Mckean, *Diffusion Processes and Their Sample paths*. Springer-Verlag, Berlin, 1965.

391

第8章 分 支 过 程

本章前 4 节对分支过程及其应用作了初步介绍. 从 8.5 节至 8.11 节进一步介绍了它的各种推广. 只有掌握了前面内容才好往后继续学习. 若只有一个学期时间, 后面的几节可以略去.

8.1 离散时间分支过程

分支过程在 2.2 节作为马尔可夫链的例子已介绍过. 马尔可夫分支过程的许多例子出现在不同的学科中. 我们列举一些比较重要的.

(a) 电子扩程器

一个电子扩程器是一个把弱电子流放大的装置. 一系列金属板放置在从某处放射出电子的路线上. 每个电子撞击在第一块金属板上, 产生随机数目的新电子, 它们又依次撞击下一个金属板, 并产生更多的电子等等. 设 X_0 是初始放射出的电子数目, X_1 是由于 X_0 个初始电子撞击在第一块金属板上产生的电子数目; 一般地, 令 X_n 是由于从第 $n-1$ 块金属板放射出的电子撞击在第 n 块金属板上所放射出的电子数目. 随机变量序列 $X_0, X_1, X_2, \cdots, X_n, \cdots$ 构成一个分支过程.

(b) 中子连锁反应

392

原子核偶然被一个中子撞击后会分裂. 裂变结果产生随机数目的新中子. 下一代中子的每一个又撞击其他某些原子核, 产生随机数目的另外一些中子, 等等. 在这种情况下, 中子的初始数目 $X_0=1$. 第一代中子包含所有那些由于初始中子撞击后发生裂变而产生的中子. 第一代中子数是随机变量 X_1. 一般地, 第 n 代的 X_n 个中子是第 $n-1$ 代的 X_{n-1} 个中子各自随机撞击原子核之后所产生的.

(c) 族姓的继承

族姓只由儿子继承. 假设每个个体以概率 p_k 有 k 个儿子. 这样, 一个个体产生出第一代, 第二代, $\cdots$, 第 n 代, $\cdots$ 的子孙后裔. 我们可以研究第 n 代后裔数目这个随机变量的分布, 或者研究绝嗣的概率. 此问题在下面分支过程的一般分析中将要涉及.

(d) 变异基因的存活

每个基因单体有可能产生出 k 个后代, $k=1,2,\cdots$, 它们是同类型的基因. 然而, 任何一个单体有可能变成不同类型或变异的基因. 这个基因可以成为多代特定

变异基因的第一代. 我们可以研究关于在原始基因群体内, 变异基因存活的可能机率.

上述各例子具有下述的构造. 令 X_0 表示初始群体的含量. 每个个体彼此独立地以概率 p_k 产生 k 个新的个体, 其中

$$p_k \geqslant 0, k = 0, 1, 2, \cdots, \quad 且 \quad \sum_{k=0}^{\infty} p_k = 1. \tag{1.1}$$

初始群体的直接后代的全体构成第一代, 其数目用 X_1 表示. 第一代每个个体又独立地产生它们的后代, 其数目也服从概率分布 (1.1). 所产生的第二代群体含量为 X_2. 一般来说, 第 n 代是由第 $n-1$ 代每个个体的后代的全体组成的, 第 $n-1$ 代的个体也各自独立地以概率 p_k 产生 k 个后代 $(k = 0, 1, 2, \cdots)$. 第 n 代个体总数为 X_n. $\{X_n\}$ 构成一列整数值随机变量, 它是一个马尔可夫链.

393

8.2 分支过程的母函数表示

我们将导出分支过程 $\{X_n\}$ 的概率母函数的几个关系式. 首先假设初始群体是由一个体组成的, 即假设 $X_0 = 1$. 显然, 对于每个 $n = 0, 1, 2, \cdots$ 我们有

$$X_{n+1} = \sum_{r=1}^{X_n} \xi_r,$$

其中 $\xi_r (r \geqslant 1)$ 是独立同分布随机变量, 其分布为

$$\Pr\{\xi_r = k\} = p_k, \quad k = 0, 1, 2, \cdots, \quad \sum_{k=0}^{\infty} p_k = 1.$$

我们定义概率母函数为

$$\varphi(s) = \sum_{k=0}^{\infty} p_k s^k$$

和

$$\varphi_n(s) = \sum_{k=0}^{\infty} \Pr\{X_n = k\} s^k, \quad 对于\ n = 0, 1, 2, \cdots.$$

显然,

$$\varphi_0(s) \equiv s \quad 且 \quad \varphi_1(s) = \varphi(s).$$

进一步有

$$\varphi_{n+1}(s) = \sum_{k=0}^{\infty} \Pr\{X_{n+1} = k\} s^k$$

$$
\begin{aligned}
&= \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \Pr\{X_{n+1}=k|X_n=j\}\Pr\{X_n=j\}s^k \\
&= \sum_{k=0}^{\infty} s^k \sum_{j=0}^{\infty} \Pr\{X_n=j\}\cdot \Pr\{\xi_1+\cdots+\xi_j=k\} \\
&= \sum_{j=0}^{\infty} \Pr\{X_n=j\}\cdot \sum_{k=0}^{\infty} \Pr\{\xi_1+\cdots+\xi_j=k\}s^k. \qquad (2.1)
\end{aligned}
$$

由于 $\xi_r(r=1,2,\cdots,j)$ 是独立同分布的随机变量, 并具有共同概率母函数 $\varphi(s)$, 因此其和 $\xi_1+\cdots+\xi_j$ 具有概率母函数 $[\varphi(s)]^j$. 这样

394

$$
\varphi_{n+1}(s) = \sum_{j=0}^{\infty} \Pr\{X_n=j\}[\varphi(s)]^j.
$$

但右边部分正是以 $\varphi(s)$ 为自度量的母函数 $\varphi_n(\cdot)$. 因此

$$
\varphi_{n+1}(s) = \varphi_n(\varphi(s)) \qquad (2.2)
$$

迭代此关系式, 可得

$$
\begin{aligned}
\varphi_{n+1}(s) &= \varphi_n(\varphi(s)) = \varphi_{n-1}(\varphi(\varphi(s))) = \varphi_{n-1}(\varphi_2(s)) \\
&= \varphi_{n-2}(\varphi_2(\varphi(s))) = \varphi_{n-2}(\varphi_3(s)).
\end{aligned}
$$

由归纳法推得, 对任何 $k=0,1,\cdots,n$,

$$
\varphi_{n+1}(s) = \varphi_{n-k}(\varphi_{k+1}(s)).
$$

特别, 令 $k=n-1$,

$$
\varphi_{n+1}(s) = \varphi(\varphi_n(s)). \qquad (2.3)
$$

如果我们改设 $X_0=i_0$(常数), 由于 $X_1=\sum_{j=1}^{i_0}\xi_j$, 因此有

$$
\varphi_0(s) \equiv s^{i_0} \quad 和 \quad \varphi_1(s) = [\varphi(s)]^{i_0}.
$$

此时我们仍有

$$
\varphi_{n+1}(s) = \varphi_n(\varphi(s)),
$$

但 (2.3) 不再成立.

借助于 (2.2) 可以计算 X_n 的期望和方差. 若无另外说明, 以后都假定 $X_0=1$. 我们还假定

$$
m = E[X_1] \quad 和 \quad \sigma^2 = \mathrm{Var}[X_1] = E[X_1^2] - [E(X_1)]^2
$$

存在且有限.

显而易见, $E[X_n]=\varphi_n'(1)$. 然后微分 (2.2), 并置 $s=1$ 得到 [由于 $\varphi(1)=1$]$\varphi_{n+1}'(1)=\varphi_n'(1)\varphi'(1)$. 迭代之, 推出

$$\varphi_{n+1}'(1)=\varphi'(1)\varphi_n'(1)=[\varphi'(1)]^2\varphi_{n-1}'(1)=[\varphi'(1)]^3\varphi_{n-2}'(1)$$

用归纳法, 得

$$\varphi_{n+1}'(1)=[\varphi'(1)]^n\varphi_1'(1)=[\varphi'(1)]^{n+1}.$$

但 $\varphi'(1)=\varphi_1'(1)=E[X_1]=m$. 因此

$$E[X_{n+1}]=m^{n+1}. \tag{2.4}$$

为计算方差 $\mathrm{Var}[X_{n+1}]$, 首先应注意

$$\varphi_n''(1)=\sum_{k=2}^{\infty}k(k-1)\Pr\{X_n=k\}=E[X_n^2]-E[X_n]=E[X_n^2]-\varphi_n'(1),$$

395

因此

$$\mathrm{Var}[X_n]=\varphi_n''(1)+\varphi_n'(1)-[\varphi_n'(1)]^2.$$

但微分 (2.3) 两次并置 $s=1$ 得

$$\varphi_{n+1}''(1)=\varphi''(1)[\varphi_n'(1)]^2+\varphi'(1)\varphi_n''(1).$$

既然 $\varphi'(1)=m$ 而 $\varphi''(1)=E[X_1^2]-E[X_1]=\sigma^2+m^2-m$, 我们有

$$\varphi_{n+1}''(1)=Mm^{2n}+m\varphi_n''(1),$$

其中 $M=\sigma^2+m^2-m$. 由归纳法得

$$\begin{aligned}\varphi_{n+1}''(1)&=M\{m^{2n}+m^{2n-1}\}+m^2\varphi_{n-1}''(1)=\cdots\\&=M\{m^{2n}+m^{2n-1}+\cdots+m^n\}.\end{aligned}$$

故

$$\begin{aligned}\mathrm{Var}[X_{n+1}]&=(\sigma^2+m^2-m)\{m^{2n}+m^{2n-1}+\cdots+m^n\}+m^{n+1}-m^{2n+2}\\&=\sigma^2\{m^{2n}+m^{2n-1}+\cdots+m^n\}\\&=\sigma^2m^n\frac{m^{n+1}-1}{m-1},\quad 若\ m\neq 1,\end{aligned}$$

并且

$$\mathrm{Var}[X_{n+1}]=(n+1)\sigma^2,\quad 若\ m=1.$$

于是我们证明了公式. $E[X_n] = m^n$ 和

$$\text{Var}[X_n] = \begin{cases} \sigma^2 m^{n-1}\dfrac{m^n - 1}{m - 1}, & 若\ m \neq 1, \\ n\sigma^2, & 若\ m = 1. \end{cases}$$

这样, 方差当 $m > 1(m < 1)$ 时是呈几何增加 (减少) 的, 当 $m = 1$ 时是线性的. 分支过程的许多结果都具有这个特征.

8.3 消失概率

我们来确定群体最终全部消失, 即绝灭的概率: $\Pr\{X_n = 0$ 对某个 $n\}$. 显然, 当 $X_n = 0$ 时, 对于所有 $k > n$ 成立 $X_k = 0$.

首先注意, 如果一个个体不产生后代的概率为 0, 即 $p_0 = 0$, 则不可能出现消失情况. 因此在研究消失概率时, 我们将假设 $0 < p_0 < 1$. 令

396

$$q_n = \Pr\{X_n = 0\} = \varphi_n(0).$$

然后由公式 (2.3) 得

$$q_{n+1} = \varphi_{n+1}(0) = \varphi(\varphi_n(0)) = \varphi(q_n). \tag{3.1}$$

由于 $\varphi(s)$ 是严格递增函数 (它是一个非负系数的幂级数且 $p_0 < 1$) 并且 $q_1 = \varphi_1(0) = p_0 > 0, q_2 = \varphi(q_1) > \varphi(0) = q_1$. 假设 $q_n > q_{n-1}$, 则 $q_{n+1} = \varphi(q_n) > \varphi(q_{n-1}) = q_n$. 这就归纳地证明了 $q_1, q_2, \cdots, q_n, \cdots$ 是单调递增序列, 并以 1 为界. 因此存在极限

$$\pi = \lim_{n\to\infty} q_n$$

并且 $0 < \pi \leqslant 1$. 既然 $\varphi(s)$ 在 $0 \leqslant s \leqslant 1$ 是连续的 [在 $s = 1$ 的连续性可由阿贝尔引理 (第 2 章引理 5.1) 推出], 在 (3.1) 中令 $n \to \infty$ 可得

$$\pi = \varphi(\pi). \tag{3.2}$$

由于 q_n 是在第 n 代或第 n 代之前消失的概率, 因此 π 是最终消失的概率, 并且 (3.2) 说明 π 是方程

$$\varphi(s) = s \tag{3.3}$$

的根.

我们现在指出 π 是 (3.3) 的最小正根. 令 s_0 是 (3.3) 的正根. 则 $q_1 = \varphi(0) < \varphi(s_0) = s_0$. 假设 $q_n < s_0$, 则由 (3.1) 有 $q_{n+1} = \varphi(q_n) < \varphi(s_0) = s_0$. 因此, 我们用归纳法证明了对于所有 n 成立 $q_n < s_0$, 由此推得 $\pi = \lim q_n \leqslant s_0$. 这就证明了 π 是 (3.3) 的最小正根.

不妨假设 $p_0 + p_1 < 1$[1]. 这时 $\varphi''(s) = \sum_{k=2}^{\infty} k(k-1)p_k S^{k-2} > 0$, $\varphi(s)$ 在 $0 <$

1. 当 $p_0 > 0, p_0 + p_1 = 1$ 时, 最终消失概率为 1.—— 译者注

$s \leqslant 1$ 是凸函数. 所以, $\varphi(s)$ 的图形与倾斜 45° 角的直线至多相交于两点. 我们知道 $\varphi(1) = 1$, 故必有交点 (1, 1). 显然可能有两种情况：若 $m = \varphi'(1) > 1$, 则图形 $\varphi(s)$ 在 $s = 1$ 处切线的斜率超过 1, 如图 8-1, 此时 $0 < \pi < 1$; 若 $m = \varphi'(1) \leqslant 1$, 则在 $s = 1$ 处切线的斜率小于或等于 1, 如图 8-2, 此时必定有 $\pi = 1$. 这样, 当 $m \leqslant 1$ 时最终消失的概率为 1, 当 $m > 1$ 时最终消失的概率小于 1. 换言之, 当且仅当每个个体后代平均数不超过 1 时, 消失是必然的.

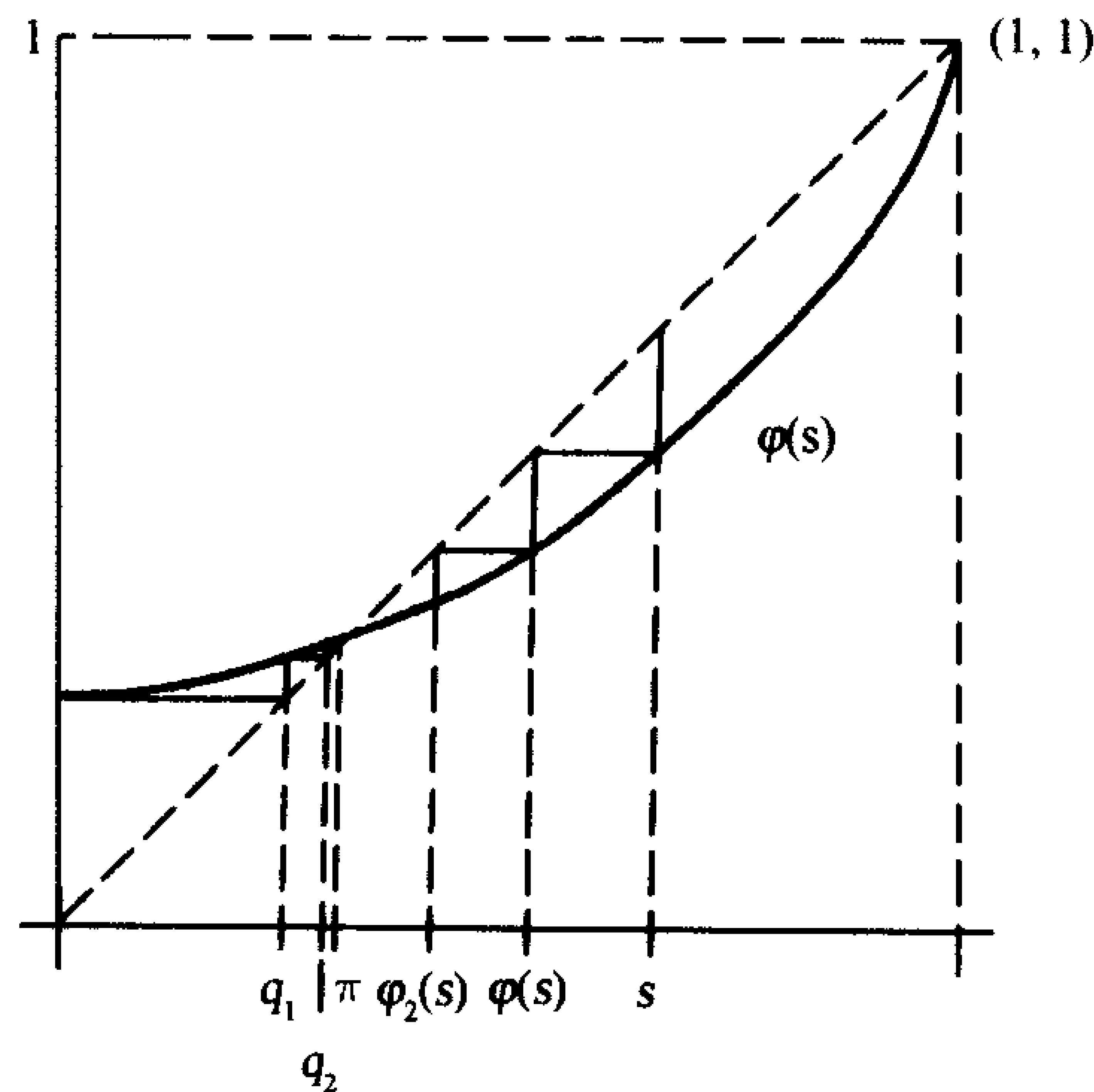


图 8-1

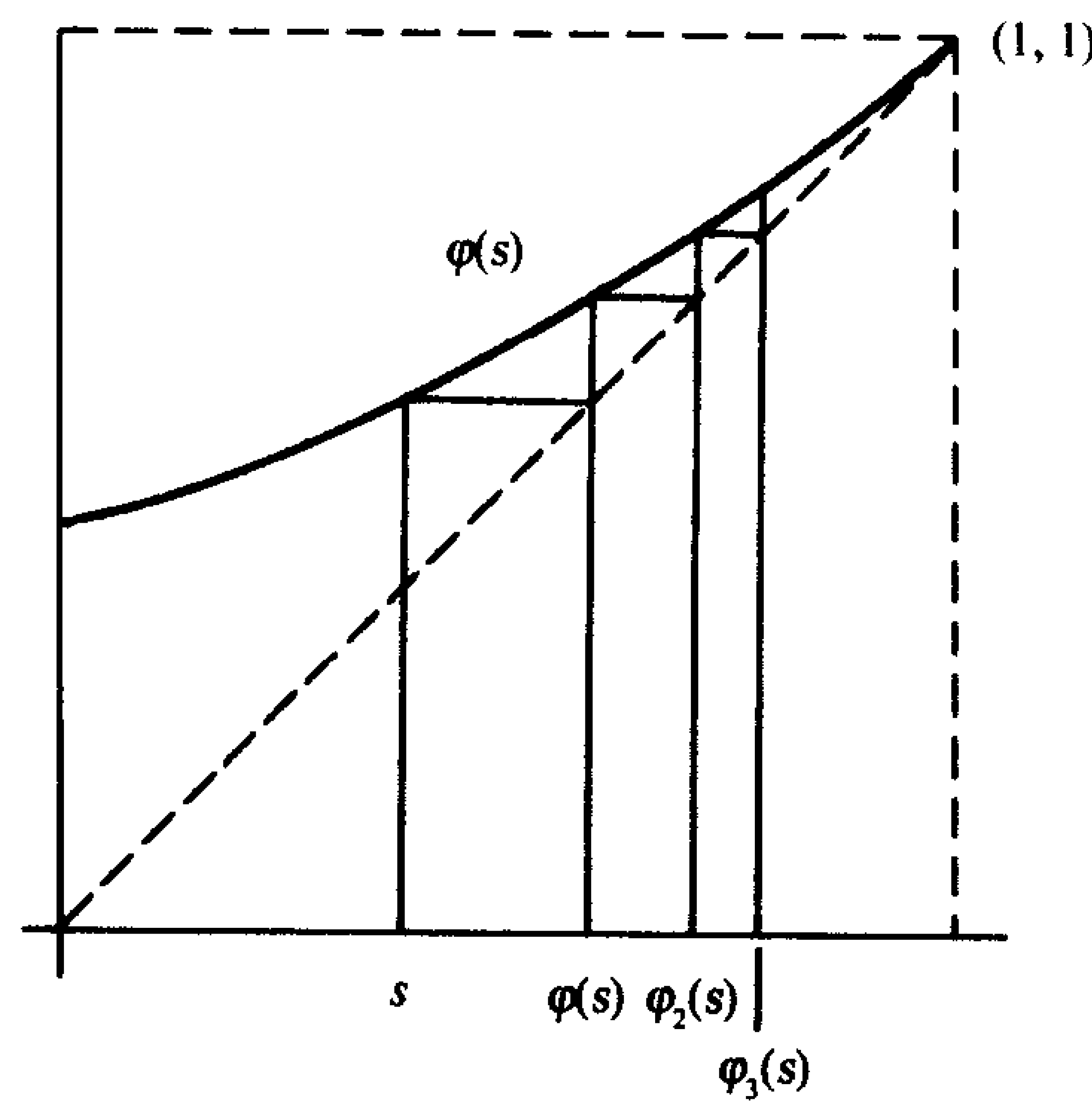


图 8-2

此外, 对于 $0 \leqslant s \leqslant \pi$, 显然有 $\varphi(s) \leqslant \varphi(\pi)$(见图 8-2). 由归纳法知, 对所有 n 有

$\varphi_n(s) \leqslant \pi (0 \leqslant s \leqslant \pi)$. 但 $\varphi_n(s) \geqslant \varphi_n(0) = q_n$, 这样 $q_n \leqslant \varphi_n(s) \leqslant \pi$. 令 $n \to \infty$, 则有

397

$$\lim_{n\to\infty} \varphi_n(s) = \pi, \quad 0 \leqslant s \leqslant \pi.$$

对于 $m > 1$ 的情况, 当 $\pi < s < 1$ 时, 我们有 $\pi < \varphi(s) < s < 1$(见图 8-1). 由归纳法知

398

$$\pi < \varphi_n(s) < \varphi_{n-1}(s) < \cdots \qquad (\pi < s < 1).$$

由此推得

$$\lim_{n\to\infty} \varphi_n(s) \geqslant \pi. \tag{3.4}$$

此极限必等于 π, 因为若 $\lim_{n\to\infty} \varphi_n(s) = \alpha > \pi$, 则 $\varphi(\alpha) < \alpha$, 但由于关系式 $\lim_{n\to\infty} \varphi_{n+1}(s) = \lim_{n\to\infty} \varphi(\varphi_n(s))$, 此时 (3.4) 式中的收敛是不可能的. 上述的分析说明

$$\lim_{n\to\infty} \varphi_n(s) = \pi \quad 对于\ 0 \leqslant s < 1.$$

$\varphi_n(s)$ 在 $0 \leqslant s < 1$ 上收敛于常数 π, 这个事实蕴涵着在级数

$$\varphi_n(s) = \sum_{k=0}^{\infty} \Pr\{X_n = k\} s^k$$

中, 当 $n \to \infty$ 时第一个系数

$$\Pr\{X_n = 0\}\ 收敛于\ \pi,$$

而所有其他系数

$$\Pr\{X_n = k\}\ 收敛于\ 0, \quad 对于\ k = 1, 2, \cdots.$$

因此, 不管 $m = EX_1 > 1$ 实际取值如何, 当 $n \to \infty$ 时第 n 代由任意有限数目个体组成的概率将趋于 0, 而消失概率趋于 π. 在这种情况下, 我们说当 $n \to \infty$ 时, 以概率 $1 - \pi$ 有 $X_n \to \infty$.

上述结果也可作为马尔可夫链一般理论的推论. 此处设马尔可夫链 $X_0, X_1, X_2, \cdots$ 具有单边吸收壁 $\{0\}$, 由于 i 和 j 是自动瞬态, 所以 $\lim_{n\to\infty} p_{ij}^n = 0, 0 \leqslant i, j < \infty$.

在本节最后, 我们指出一个有趣性质. 当已知 X_n 时, X_{n+r}(r 是正整数) 的条件期望是 $m^r. X_n$, 即 $E(X_{n+r}|X_n) = m^r X_n$. 为证明这点, 首先考虑 $r = 1$ 的情况:

$$E\{X_{n+1}|X_n\} = E\left\{\sum_{j=1}^{X_n} \xi_j | X_n\right\} = X_n E\xi_j = mX_n.$$

现在假设所述关系式对于 r 成立, 继而证明对于 $r+1$ 也成立.

$$\begin{aligned}
&E\{X_{n+r+1}|X_n\}\\
&\quad=E\{E[X_{n+r+1}|X_{n+r},X_{n+r-1},\cdots,X_n]|X_n\}\\
&\quad=E\{E[X_{n+r+1}|X_{n+r}]|X_n\},
\end{aligned}$$

其中我们使用了 $\{X_n\}$ 的马尔可夫性质. 但

$$E[X_{n+r+1}|X_{n+r}]=X_{n+r}\cdot m,$$

由归纳假设得

$$E\{mX_{n+r}|X_n\}=m^{r+1}X_n$$

这样, 我们证明了

$$E\{X_{n+r}|X_n\}=X_nm^r\quad 对于\ r=0,1,2,\cdots,n=0,1,2,\cdots \tag{3.5}$$

399

现在考虑随机变量

$$W_n=\frac{X_n}{m^n},\quad n=0,1,2,\cdots.$$

根据 (3.5), 我们有

$$E\{W_{n+r}|W_n\}=\frac{1}{m^{n+r}}E\{X_{n+r}|X_n\}=\frac{1}{m^{n+r}}\cdot X_n\cdot m^r=W_n.$$

对于 $r,n=0,1,2,\cdots$, 我们有

$$E\{W_{n+r}|W_n,W_{n-1},\cdots,W_1,W_0\}=W_n, \tag{3.6}$$

由此说明 $\{W_n\}_{n=0}^{\infty}$ 是一个鞅.

8.4 例 子

(i) 设 $\varphi(s)=p_0+p_1s, 0<p_0<1$. 对应的分支过程是纯灭过程. 在各个周期, 每个个体以概率 p_0 独立地消亡, 以概率 $1-p_0=p_1$ 存活.

(ii) 设 $\varphi(s)=p_0+p_2s^2(0<p_0<1,p_0+p_2=1)$. 这是对应于下面分支过程的概率母函数, 在此过程中每一代的每个个体或者消亡或者被两个后代所代替.

(iii) 考虑这样一个例子, 每个个体分别以概率 p 和 q 产生 N 个或 0 个后代. 这样 $p_0=q, p_N=p$, 并且对于 $k\neq 0,N,p_k=0$. 则

$$\varphi(s)=q+ps^N. \tag{4.1}$$

(iv) 每个个体有 k 个后代, 其中 k 是具有参数为 N 和 p 的二项概率分布, 则

$$\varphi(s)=(q+ps)^N. \tag{4.2}$$

(v) 关于本章开头描述的例 (d), 常常假设变异基因具有 k 个直接后代 ($k=0,1,2,\cdots$) 的概率服从均值为 $\lambda=1$ 的泊松分布, 则 $\varphi(s)=\mathrm{e}^{s-1}$ 且 $\pi=1$.

选择此分布的理由如下. 在许多群体中, 虽有大量受精卵 (受精蛋) 生成, 但仅有少部分发育成熟. 受精事件和不同受精卵的成熟服从独立二项分布律. 试验的次数 (即受精卵的数目) 是非常大的, 故实际成熟后代的数目遵循泊松分布. 在罕见变

400

异基因群体增长问题中, 此模型是十分适当的. 如果变异基因具有某种生物学上的优势 (或劣势), 则其概率分布可取为均值 $\lambda>1$(或 $\lambda<1$) 的泊松分布. 此时,

$$\varphi(s)=\mathrm{e}^{\lambda(s-1)} \tag{4.3}$$

并且当且仅当 $\lambda>1$ 时有 $0<\pi<1$.

在一个变异基因的异源群体中, 我们可以假设后代数目的概率分布是泊松分布, 但其均值也是一个随机变量.

例如, 我们可能有一个很大地域, 在每一块子区域发生分支过程, 它由参数为 λ 的泊松分布的概率母函数所刻划. 此外, 我们假设 λ 的值依赖于每块子区域, 并且在整块地域上, 其分布服从 Γ 分布. 从形式上, 我们假设一个变异基因恰有 k 个直接后代的概率由下式给出:

$$p_k=\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!},\quad k=0,1,2,\cdots,$$

其中 λ 本身是服从 Γ 分布的随机变量, 其密度函数为

$$f(\lambda)=\begin{cases}\dfrac{(q/p)^\alpha\lambda^{\alpha-1}}{\Gamma(\alpha)}\exp\left(-\dfrac{q}{p}\lambda\right), & \text{对于}\lambda\geqslant 0,\\ 0, & \text{其他},\end{cases}$$

其中 q,p,α 是正常数, 并且 $q+p=1$. 一个个体具有 k 个直接后代的概率是

$$\Pr\{\xi=k\}=\int_0^\infty \Pr\{\xi=k|\lambda\}f(\lambda)\mathrm{d}\lambda,$$

上式是关于参数 λ 求平均的. 因此, 其母函数是

$$\begin{aligned}\varphi(s)&=\sum_{k=0}^\infty \Pr\{\xi=k\}s^k\\ &=\sum_{k=0}^\infty\int_0^\infty \exp(-\lambda)\frac{\lambda^k}{k!}\frac{(q/p)^\alpha\lambda^{\alpha-1}}{\Gamma(\alpha)}\exp\left(-\frac{q}{p}\lambda\right)\mathrm{d}\lambda\cdot s^k\end{aligned}$$

$$
\begin{aligned}
&= \int_0^\infty \exp(-\lambda) \frac{(q/p)^\alpha \lambda^{\alpha-1}}{\Gamma(\alpha)} \exp\left(-\frac{q}{p}\lambda\right) \left(\sum_{k=0}^\infty \frac{(\lambda s)^k}{k!}\right) \mathrm{d}\lambda \\
&= \int_0^\infty \exp\left\{-\left(\frac{q}{p}+1-s\right)\lambda\right\} \frac{(q|p)^\alpha \lambda^{\alpha-1}}{\Gamma(\alpha)} \mathrm{d}\lambda \\
&= \left(\frac{q/p}{(q/p)+1-s}\right)^\alpha = \left(\frac{q}{1-ps}\right)^\alpha .
\end{aligned}
$$

401

我们知道这是负二项分布的母函数.

(vi) 在例 (ii)~(iv) 中, 我们并不知道第 n 代概率母函数 $\varphi_n(s)$ 的简短形式的表达式. 经过全面分析, 下面剖析的例子是正确的. 特别地, 我们可求出第 n 代概率母函数的表达式. 设

$$p_k = bc^{k-1}, \quad k = 1, 2, \cdots,$$

并且

$$p_0 = 1 - \sum_{k=1}^\infty p_k,$$

其中 $b, c > 0$ 和 $b + c < 1$. 我们有

$$p_0 = 1 - b\sum_{k=1}^\infty c^{k-1} = 1 - \frac{b}{1-c} = \frac{1-b-c}{1-c}$$

而对应的概率母函数是

$$\varphi(s) = 1 - \frac{b}{1-c} + bs\sum_{k=1}^\infty (cs)^{k-1} = \frac{1-(b+c)}{1-c} + \frac{bs}{1-cs}. \tag{4.4}$$

注意, $\varphi(s)$ 是如下线性分式变换的一种形式:

$$f(s) = \frac{\alpha + \beta s}{\gamma + \delta s}, \quad \alpha\sigma - \beta\gamma \neq 0. \tag{4.5}$$

我们列出后面要用到的线性分式变换的几个初等性质:

(i) 迭代线性分式变换仍然是线性分式变换, 因为如果 $f(s)$ 是由 (4.5) 定义的, 通过简单的代数运算得到

$$f(f(s)) = \frac{\alpha(\gamma+\beta) + (\alpha\delta + \beta^2)s}{\alpha\delta + \gamma^2 + \delta(\gamma+\beta)s}.$$

(ii) 方程 $f(s) = s$ 总存在两个有限的解 (可能相等). 此解称为 $f(\cdot)$ 的固定点. 如果 $f(s)$ 是一个概率母函数, 则 $S_1 = 1$ 必是其中一个固定点, 并且另外一个固定点 s_0 小于、等于或大于 1, 分别对应于 $f'(1)$ 大于、等于或小于 1.

对于由 (4.4) 给出的母函数, 可以通过代数方法直接验证第二个固定点是

$$S_0 = \frac{1-b-c}{c(1-c)},$$

402 其中 $c > 0, b + c < 1$.

(iii) 对于任何两个点 $S_i, i = 0, 1$, 显而易见有

$$\frac{f(s)-f(s_i)}{s-s_i} = \frac{\gamma\beta - \alpha\delta}{(\gamma+\delta s)(\gamma+\delta s_i)},$$

因此,

$$\frac{f(s)-f(s_0)}{f(s)-f(s_1)} = \left(\frac{\gamma+\delta s_1}{\gamma+\delta s_0}\right)\left(\frac{s-s_0}{s-s_1}\right). \tag{4.6}$$

如果令 s_0 和 s_1 是 $f(\cdot)$ 的两个 (不相等的) 固定点, 并记 $w = f(s)$, (4.6) 变成

$$\frac{w-s_0}{w-s_1} = \kappa\frac{s-s_0}{s-s_1}, \tag{4.7}$$

其中, 通过设置 $s = 0$ k 可以由 (4.6) 及 (4.5) 得到. 后者计算较为简便.

利用 (4.7), 我们易于得到 $f(s)$ 的迭代值 $f_n(s) = W_n$:

$$\frac{w_2-s_0}{w_2-s_1} = \kappa\frac{w_1-s_0}{w_1-s_1} = \kappa\left(\kappa\frac{s-s_0}{s-s_1}\right),$$

并且, 一般地有

$$\frac{w_n-s_0}{w_n-s_1} = \kappa^n\frac{s-s_0}{s-s_1}. \tag{4.8}$$

对于由 (4.4) 给出的几何分布的母函数, 注意到固定点是 $s_0 = (1-b-c)/c(1-c)$ 和 $s_1 = 1$, 我们得

$$\kappa = \frac{(1-c)^2}{b} = \frac{1}{m},$$

其中 m 是此几何分布的均值. 对于 $m \neq 1$, 两个固定点 s_0 和 1 是不同的; 因此在 (4.8) 中解 W_n 得:

$$w_n = \frac{s_0 - (1/m^n)[(s-s_0)/(s-1)]}{1-(1/m^n)[(s-s_0)/(s-1)]}, \quad m \neq 1, \tag{4.9}$$

此式可改写为

$$\varphi_n(s) = 1 - m^n\left(\frac{1-s_0}{m^n-s_0}\right) + \frac{m^n[(1-s_0)/(m^n-s_0)]^2 s}{1-[(m^n-1)/(m^n-s_0)]s}. \tag{4.10}$$

因此, 第 n 代消失概率是

403

$$\Pr\{X_n = 0\} = \varphi_n(0) = 1 - m^n\left(\frac{1-s_0}{m^n-s_0}\right).$$

注意, 若令 $n\to\infty$, 这个表达式当 $m>1$ 收敛于 s_0, 当 $m<1$ 收敛于 1. 将 (4.10) 简单展开为 s 的幂级数, 便得到群体第 n 代含量为 k 的概率, $\Pr\{X_n=k\}, k=1,2,\cdots$. 如果我们定义消失时间 T 为使得 $X_n=0$ 的 n 的最小值, 即首达状态 0 的时刻, 则

$$\Pr\{T\leqslant n\}=\Pr\{X_n=0\}=\varphi_n(0),$$

因此,

$$\Pr\{T=n\}=\Pr\{T\leqslant n\}-\Pr\{T\leqslant n-1\}=\varphi_n(0)-\varphi_{n-1}(0).$$

在 $m\neq 1$ 的情况下, 我们有

$$\begin{aligned}\Pr\{T=n\}&=1-m^n\left(\frac{1-s_0}{m^n-s_0}\right)-1+m^{n-1}\left(\frac{1-s_0}{m^{n-1}-s_0}\right)\\&=m^{n-1}s_0\frac{(m-1)(1-s_0)}{(m^n-s_0)(m^{n-1}-s_0)},\quad 对于\ n=1,2,\cdots.\end{aligned}$$

如果 $m=1$, 则 $b=(1-c)^2$ 并且方程 $\varphi(s)=s$ 具有重根 $s=1$, 而没有其他的根. 此时

$$\varphi(s)=c+\frac{(1-c)^2s}{1-cs}=\frac{c-(2c-1)s}{1-cs},$$

并且

$$\varphi_2(s)=\varphi(\varphi(s))=\frac{c-(2c-1)[(c-(2c-1)s)/(1-cs)]}{1-c[(c-(2c-1)s)/(1-cs)]}=\frac{2c-(3c-1)s}{1+c-2cs},$$

由归纳法得

$$\varphi_n(s)=\frac{nc-[(n+1)c-1]s}{1+(n-1)c-ncs}. \tag{4.11}$$

在 $m=1$ 的情况下, 消失概率为

$$\Pr\{X_n=0\}=\varphi_n(0)=\frac{nc}{1+(n-1)c},\quad 对于\ n=1,2,\cdots;$$

且消失时间 T 的分布为

$$\Pr\{T=n\}=\frac{nc}{1+(n-1)c}-\frac{(n-1)c}{1+(n-2)c}=\frac{c(1-c)}{[1+(n-1)c][1+(n-2)c]}.$$

8.5 二维分支过程

我们把前面结果推广到二维情况. 考虑某生物群体或某研究对象全体, 其中有两种类型的个体可区分. 每一种类型个体都可能独立产生两种类型的后代个体. 设 404

U_n 和 V_n 分别是第 n 代 I 型和 II 型个体的数目. 我们有如下表示式:

$$U_{n+1} = \sum_{j=1}^{U_n} \xi_j^{(1)} + \sum_{j=1}^{V_n} \xi_j^{(2)},$$

$$V_{n+1} = \sum_{j=1}^{U_n} \zeta_j^{(1)} + \sum_{j=1}^{V_n} \zeta_j^{(2)},$$

此处 $(\xi_j^{(i)}, \zeta_j^{(i)})$ 是独立同分布的随机向量, 其分布律为

$$\Pr\{\xi_j^{(i)} = k, \zeta_j^{(i)} = l\} = p_i(k,l), k, l = 0, 1, 2, \cdots, j = 1, 2, \cdots, i = 1, 2.$$

这里, $pi(k,l) \geqslant 0$ 且 $\sum_{k,l=0}^{\infty} p_i(k,l) = 1, i = 1, 2.$

换言之, $p_1(k,l)$ 和 $p_2(k,l)$ 分别是 I 型和 II 型的单个体产生 $k+l$ 个直接后代并且 k 个是 I 型而 l 个是 II 型的概率.

假设过程初始时刻仅有一个个体, 即或者

$$U_0 = 1 \quad \text{和} \quad V_0 = 0 \tag{5.1}$$

或者

$$U_0 = 0 \quad \text{和} \quad V_0 = 1. \tag{5.2}$$

下面我们引入一对二维概率母函数

$$\varphi^{(i)}(s,t) = \sum_{k,l=0}^{\infty} p_i(k,l) s^k \cdot t^l, \quad i = 1, 2,$$

即,

$$\varphi_n^{(1)}(s,t) = \sum_{k,l=0}^{\infty} \Pr\{U_n = k, V_n = l | U_0 = 1, V_0 = 0\} s^k \cdot t^l,$$

$$\varphi_n^{(2)}(s,t) = \sum_{k,l=0}^{\infty} \Pr\{U_n = k, V_n = l | U_0 = 0, V_0 = 1\} s^k \cdot t^l.$$

显然, (5.1) 的母函数是

$$\varphi_0^{(1)}(s,t) \equiv s,$$

及 (5.2) 的母函数是

$$\varphi_0^{(2)}(s,t) \equiv t.$$

而且,

$$\varphi_1^{(i)}(s,t) = \varphi^{(i)}(s,t), \quad i = 1,2.$$

405

类似于一维情形, 可以证明对于 $i=1,2$ 和 $n,m=0,1,2,\cdots$, 成立

$$\varphi_{n+m}^{(i)}(s,t) = \varphi_m^{(i)}(\varphi_n^{(1)}(s,t), \varphi_n^{(2)}(s,t)). \tag{5.3}$$

此式与 (2.3) 二维情形相当.

为推广公式 (3.5), 我们引入下面记号. 设 $\boldsymbol{X}_n = (U_n, V_n)$ 是具有分量 U_n 和 V_n 的二维向量. 记

$$\begin{aligned}
m_{11} &= E\{U_1|U_0=1, V_0=0\} = E\xi^{(1)},\\
m_{12} &= E\{V_1|U_0=1, V_0=0\} = E\zeta^{(1)},\\
m_{21} &= E\{U_1|U_0=0, V_0=1\} = E\xi^{(2)},\\
m_{22} &= E\{V_1|U_0=0, V_0=1\} = E\zeta^{(2)}.
\end{aligned}$$

引入期望矩阵

$$\boldsymbol{M} = \left\| \begin{matrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{matrix} \right\|.$$

这样, m_{11} 和 m_{12} 分别是由单个 I 型个体产生的 I 型和 II 型后代的期望数. 作为式 (3.5) 的推广, 我们有矩阵恒等式

$$E[\boldsymbol{X}_{n+r}|\boldsymbol{X}_n] = \boldsymbol{X}_n \boldsymbol{M}^r, \quad 对于\ r,n=0,1,2,\cdots. \tag{5.4}$$

上式当 $r=1$ 时的证明可直接进行如下:

$$\begin{aligned}
E[\boldsymbol{X}_{n+1}|\boldsymbol{X}_n] &= \left(E\left[\sum_{j=1}^{U_n}\xi_j^{(1)} + \sum_{j=1}^{V_n}\xi_j^{(2)}|(U_n,V_n)\right], E\left[\sum_{j=1}^{U_n}\zeta_j^{(1)} + \sum_{j=1}^{V_n}\zeta_j^{(2)}|(U_n,V_n)\right]\right)\\
&= (m_{11}U_n + m_{21}V_n, m_{12}U_n + m_{22}V_n)\\
&= (U_n, V_n)\left\| \begin{matrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{matrix} \right\|\\
&= \boldsymbol{X}_n \cdot \boldsymbol{M}.
\end{aligned}$$

现假设式 (5.4) 关系式对于 r 成立, 继而证它对 $r+1$ 也成立. 由 $\{\boldsymbol{X}_n\}$ 的马尔可夫性, 我们有

$$\begin{aligned}
E[\boldsymbol{X}_{n+r+1}|\boldsymbol{X}_n] &= E\{E[\boldsymbol{X}_{n+r+1}|\boldsymbol{X}_{n+r},\cdots,\boldsymbol{X}_n]|\boldsymbol{X}_n\}\\
&= E\{E[\boldsymbol{X}_{n+r+1}|\boldsymbol{X}_{n+r}]|\boldsymbol{X}_n\} = E\{\boldsymbol{X}_{n+r}\boldsymbol{M}|\boldsymbol{X}_n\}\\
&= E\{\boldsymbol{X}_{n+r}|\boldsymbol{X}_n\}\cdot \boldsymbol{M} \quad (由归纳假设知)\\
&= \boldsymbol{X}_n\boldsymbol{M}^{r+1}.
\end{aligned}$$

406

这就归纳证明了 (5.4).

对于二维分支过程, 我们也可引入消失概率：

$$\pi^{(1)} = \Pr\{U_n = V_n = 0 \quad \text{对某个 } n | U_0 = 1, V_0 = 0\},$$

$$\pi^{(2)} = \Pr\{U_n = V_n = 0 \quad \text{对某个 } n | U_0 = 0, V_0 = 1\}.$$

一维情形有关结果可以推广到二维情况, 此时应注意在二维情况下期望 m 所起的作用由期望矩阵 $\boldsymbol{M}$ 的最大特征值 p 代替.

建议读者可先阅读附录有关内容, 特别是 Frobenius 定理, 该定理指出, 如果 $\boldsymbol{M}$ 的每个元素都是正的 (用符号表示为 $\boldsymbol{M} \gg 0$), 则其最大特征值是正的且为实的. 此特征值记为 $p(\boldsymbol{M}) = p$.

为方便表示, 我们引用如下向量记号：

$$\begin{aligned} \boldsymbol{u} &= (s, t), \\ \boldsymbol{\phi}(\boldsymbol{u}) &= (\varphi^{(1)}(s,t), \varphi^{(2)}(s,t)), \\ \boldsymbol{\phi_n}(\boldsymbol{u}) &= (\varphi_n^{(1)}(s,t), \varphi_n^{(2)}(s,t)), \\ \boldsymbol{\pi} &= (\pi^{(1)}, \pi^{(2)}), \\ \mathbf{1} &= (1,1). \end{aligned}$$

于是我们可叙述

定理 5.1 假设 $\boldsymbol{\phi}(\boldsymbol{u})$ 的分量是 s 和 t 的非线性函数且 $\boldsymbol{M} \gg 0$(M 中每个元素都是正的), 若 $\boldsymbol{M}$ 的最大特征值 p 不超过 1 则 $\boldsymbol{\pi} = \mathbf{1}$, 若 $p > 1$ 则 $\boldsymbol{\pi} \ll \mathbf{1}$. (注：记号 $\boldsymbol{u} \ll \boldsymbol{v} (u \leqslant v)$ 表示 $\boldsymbol{v} - \boldsymbol{u}$ 的每一个元素都是正的, $\boldsymbol{u} \leqslant \boldsymbol{v}$ 表示 $\boldsymbol{v} - \boldsymbol{u}$ 的每一个元素都是非负的.) 当 $p > 1$ 时, $\boldsymbol{\pi}$ 是下面方程最小非负解：

$$\boldsymbol{u} = \boldsymbol{\phi}(\boldsymbol{u}), \quad \boldsymbol{u} \ll \mathbf{1}. \tag{5.5}$$

证明 先考虑情况 $p \leqslant 1$. 根据马尔可夫链的一般理论, 如果链只有一个吸收状态, 从某状态出发可以到达吸收状态, 则这个出发的状态便是瞬态. 二维过程 $\boldsymbol{X_n} = (U_n, V_n)$ 正是这样一个过程：原点是仅有的吸收状态, 并且可从所有其他状态到达吸收状态. 这是由于 $\boldsymbol{\phi}(\boldsymbol{u})$ 没有线性分量并且 $p \leqslant 1$. 这样, 除了原点外, 每一个状态都是瞬态. 若 $|\boldsymbol{X}_n| = \boldsymbol{U}_n + \boldsymbol{V}_n$, 则对任意正数 N 有

$$\Pr\{0 < |\boldsymbol{X}_n| < N \text{ 有无限多 } n\} = 0$$

(参见第 2 章定理 7.1.) 上式意味着

407

$$\Pr\{|\boldsymbol{X}_n| \to 0\} + \Pr\{|\boldsymbol{X}_n| \to \infty\} = 1.$$

由公式 (5.4) 有 $E[\boldsymbol{X}_n | \boldsymbol{X}_0] = \boldsymbol{X}_0 \boldsymbol{M}^n$. 而附录定理 2.3 断言, 当 $n \to \infty$ 时, $(1/p^n)\boldsymbol{M}^n$ 的各分量收敛. 因此, 在 $p \leqslant 1$ 的情况下, 当 $n \to \infty$ 时, $E[\boldsymbol{X}_n | \boldsymbol{X}_0]$ 各分量仍有界,

由此推得事件 $|\boldsymbol{X}_n| \to \infty$ 出现的概率为 0. 所以 $\Pr\{|\boldsymbol{X}_n| \to 0\} = 1$, 或等价地, 当 $n \to \infty$ 时以概率 1 有 $U_n \to 0$ 和 $V_n \to 0$ 成立. 这样, 若 $p \leqslant 1$, 则

$$\pi^{(1)} = \pi^{(2)} = 1.$$

下一步考虑情况 $p > 1$. 由公式 (5.3), 当 $s = t = 0$ 时有

$$\varphi_{n+1}^{(i)}(0,0) = \varphi^{(i)}(\varphi_n^{(1)}(0,0), \varphi_n^{(2)}(0,0)), \quad i = 1, 2. \tag{5.6}$$

令

$$\begin{aligned} q_n^{(1)} &= \varphi_n^{(1)}(0,0) = \Pr\{U_n = V_n = 0 | U_0 = 1, V_0 = 0\}, \\ q_n^{(2)} &= \varphi_n^{(2)}(0,0) = \Pr\{U_n = V_n = 0 | U_0 = 0, V_0 = 1\}. \end{aligned}$$

则 (5.6) 可改写为

$$q_{n+1}^{(i)} = \varphi^{(i)}(q_n^{(1)}, q_n^{(2)}), \quad i = 1, 2. \tag{5.7}$$

由于 $\varphi^{(i)}(s,t)$ 关于变量 s 和 t 是递增的, 且若 s, t 均增加且严格递增时, 由 $q_1^{(i)} = \varphi_1^{(i)}(0,0) > 0, i = 1, 2$, 我们易得

$$q_2^{(i)} = \varphi^{(i)}(q_1^{(1)}, q_1^{(2)}) > \varphi^{(i)}(0,0) = q_1^{(i)}, \quad i = 1, 2.$$

由归纳可得

$$q_{n+1}^{(i)} = \varphi^{(i)}(q_n^{(1)}, q_n^{(2)}) > \varphi^{(i)}(q_{n-1}^{(1)}, q_{n-1}^{(2)}) = q_n^{(i)}, \quad i = 1, 2.$$

因此, 对于每个 $i = 1, 2, q_n^{(i)}$ 构成一单调增加且上界为 1 的序列, 且

$$\lim_{n \to \infty} q_n^{(i)} = \pi^{(i)} \leqslant 1, \quad i = 1, 2.$$

在 (5.7) 中令 $n \to \infty$, 则得

$$\pi^{(i)} = \varphi^{(i)}(\pi^{(1)}, \pi^{(2)}), \quad i = 1, 2,$$

或用向量记号: $\boldsymbol{\pi} = \boldsymbol{\phi}(\boldsymbol{\pi})$. 现在我们证 $\boldsymbol{\pi} \ll \boldsymbol{1}$, 且在所述条件之下, 它是 (5.5) 唯一的解. 根据泰勒定理在点 (1.1) 处展开 $\varphi_n^{(i)}(\cdot,\cdot)$, 有

$$\begin{aligned} \varphi_n^{(j)}(1-s, 1-t) = &\varphi_n^{(i)}(1,1) - \left(\left. \frac{\partial \varphi_n^{(i)}(s,t)}{\partial s} \right|_{s=t=1} \right) s \\ &- \left(\left. \frac{\partial \varphi_n^{(i)}(s,t)}{\partial t} \right|_{s=t=1} \right) t + o(|s| + |t|), \end{aligned} \tag{5.8}$$

408

上式对于 $|1-s| \leqslant 1, |1-t| \leqslant 1$ 及 s,t 足够小时正确. 其中符号 $o(\cdot)$ 表示当 $|s|+|t| \to 0$ 时 $[o(|s|+|t|)]/(|s|+|t|) \to 0$. 此外,

$$\left.\frac{\partial \varphi_n^{(1)}(s,t)}{\partial s}\right|_{s=t=1} = E[U_n|U_0=1, V_0=0] = m_{11}^{(n)},$$
$$\left.\frac{\partial \varphi_n^{(1)}(s,t)}{\partial t}\right|_{s=t=1} = E[V_n|U_0=1, V_0=0] = m_{12}^{(n)},$$
$$\left.\frac{\partial \varphi_n^{(2)}(s,t)}{\partial s}\right|_{s=t=1} = E[U_n|U_0=0, V_0=1] = m_{21}^{(n)},$$
$$\left.\frac{\partial \varphi_n^{(2)}(s,t)}{\partial t}\right|_{s=t=1} = E[V_n|U_0=0, V_0=1] = m_{22}^{(n)},$$

我们可把 (5.8) 写为向量形式

$$\boldsymbol{\varphi}_n(\mathbf{1}-\boldsymbol{u}) = \mathbf{1} - \boldsymbol{M}^{(n)}\boldsymbol{u} + o(|s|+|t|), \tag{5.9}$$

其中

$$\boldsymbol{M}^{(n)} = \left\| \begin{matrix} m_{11}^{(n)}, & m_{12}^{(n)} \\ m_{21}^{(n)}, & m_{22}^{(n)} \end{matrix} \right\|$$

且 $|\boldsymbol{u}| < \varepsilon$. 显然, 由关系式 $E[\boldsymbol{X}_n|\boldsymbol{X}_0] = \boldsymbol{X}_0\boldsymbol{M}^n$ 可得 $\boldsymbol{M}^{(n)} = \boldsymbol{M}^n$. 设向量 $\boldsymbol{v} = (v_1, v_2)$ 的绝对值定义为各坐标分量绝对值之和: $|\boldsymbol{v}| = |v_1| + |v_2|$. 我们证明, 倘若 $\boldsymbol{u} \geqslant o$, 对于足够大的 n, 有

$$|\boldsymbol{M}^n\boldsymbol{u}| > 2|\boldsymbol{u}|, \quad \boldsymbol{u} = (s,t). \tag{5.10}$$

事实上, 根据附录的定理 2.3, 我们有

$$\boldsymbol{M}^n\boldsymbol{u} = \rho^n \left\| \begin{matrix} x_1^0 y_1^0 & x_1^0 y_2^0 \\ x_2^0 y_1^0 & x_2^0 y_2^0 \end{matrix} \right\| \cdot \boldsymbol{u} + o(\rho^n)\boldsymbol{u},$$

其中 ρ 是 $\boldsymbol{M}$ 的最大特征值, 并且 $\boldsymbol{x}^0 = (x_1^0, x_2^0)$ 和 $\boldsymbol{y}^0 = (y_1^0, y_2^0)$ 表示是唯一的 (可相差一个常数因子) 左和右的正规化正特征向量使得 $x_1^0 y_1^0 + x_2^0 y_2^0 = 1$. 至于 $o(\rho^n)$ 的意义是通常意义下的推广, 即当 $n \to \infty$ 时, $(o(\rho^n))/\rho^n$ 的分量都趋于零. $o(\rho^n)$ 不依赖于 $\boldsymbol{u}$, 它表示当 $\boldsymbol{M}^n/\rho^n$ 收敛到极限时的误差项. 我们改写上述表达式为

$$\boldsymbol{M}^n\boldsymbol{u} = \rho^n(y_1^0 s + y_2^0 t)\boldsymbol{x}^0 + o(\rho^n)\boldsymbol{u}, \quad \boldsymbol{u} = (s,t),$$

且若 $\boldsymbol{u} \geqslant 0$, 我们得到估计式

409

$$|\boldsymbol{M}^n\boldsymbol{u}| \geqslant \rho^n[x_1^0 + x_2^0]\min(y_1^0, y_2^0) \cdot |\boldsymbol{u}| + o(\rho^n)|\boldsymbol{u}|.$$

由于 $\rho > 1$, 选择 n 足够大即得 (5.10) 成立. 综合 (5.9) 和 (5.10), 若 $\mathbf{1} \geqslant \boldsymbol{u} \geqslant 0, |\boldsymbol{u}|$ 充分小以及 n 足够大时, 例如 $n \geqslant n_0$, 我们导出

$$|\mathbf{1} - \boldsymbol{\varphi}_n(\mathbf{1} - \boldsymbol{u})| > 2|\boldsymbol{u}|,$$

令 $\boldsymbol{v} = \mathbf{1} - \boldsymbol{u}$, 则对于 $0 \leqslant \boldsymbol{v} \leqslant \mathbf{1}$, $|\mathbf{1} - \boldsymbol{v}| < \varepsilon$ 和 $n \geqslant n_0$, 有

$$|\mathbf{1} - \boldsymbol{\varphi}_n(\boldsymbol{v})| > 2|\mathbf{1} - \boldsymbol{v}|. \tag{5.11}$$

现在我们利用 (5.11) 来证明 $\boldsymbol{\pi} \ll \mathbf{1}$. 假设 $\boldsymbol{\pi} = \mathbf{1}$, 即 $\pi^{(i)} = 1, i = 1, 2$. 此时 $q_n^{(i)} = \varphi_n^{(i)}(0) \geqslant 0$ 当 $n \to \infty$ 时逼近于 1. 由 (5.3), 我们有

$$\boldsymbol{\phi}_{n+N}(0) = \boldsymbol{\phi}_n(\boldsymbol{\phi}_N(0)).$$

在 (5.11) 中令 $\boldsymbol{v} = \boldsymbol{\phi}_N(0)$, 知

$$|\mathbf{1} - \boldsymbol{\phi}_{n+N}(\mathbf{0})| = |\mathbf{1} - \boldsymbol{\phi}_n(\boldsymbol{\phi}_N(\mathbf{0}))| > 2|\mathbf{1} - \boldsymbol{\phi}_N(\mathbf{0})| \tag{5.12}$$

仅当 $|\mathbf{1} - \boldsymbol{\phi}_N(\mathbf{0})| < \varepsilon$ 时成立, 后者当 N 足够大时能够实现. 然而, 关系式 (5.12) 与假设当 $n \to \infty$ 时 $\varphi_n^{(i)}(0)$ 趋于 1 矛盾. 这样 $\pi^{(1)} = \pi^{(2)} = 1$ 是不可能的. 现在假设 $\pi^{(1)} < 1$ 和 $\pi^{(2)} = 1$. 那么

$$\pi^{(1)} = \varphi^{(1)}(\pi^{(1)}, 1)$$

且

$$1 = \pi^{(2)} = \varphi^{(2)}(\pi^{(1)}, 1).$$

于是, 我们有

$$\varphi^{(2)}(1,1) = 1 \quad 和 \quad \varphi^{(2)}(\pi^{(1)}, 1) = 1,$$

其中 $\pi^{(1)} < 1$. 由于 $\varphi^{(2)}(s,t)$ 关于 s 是单调的, $\varphi^{(2)}(s,1)$ 在区间 $\pi^{(1)} \leqslant s \leqslant 1$ 必须是常数; 所以

$$\frac{\partial \varphi^{(2)}(s,1)}{\partial s} = 0, \quad 当\ \pi^{(1)} \leqslant s \leqslant 1$$

而且

$$m_{21}^{(2)} = \left.\frac{\partial \varphi^{(2)}(s,t)}{\partial s}\right|_{s=t=1} = 0.$$

显然, 这与我们假设 $\boldsymbol{M} \gg 0$ 矛盾. 用类似方法可推得 $\pi^{(1)} = 1, \pi^{(2)} < 1$ 也是不可能的. 这样即证得 $\boldsymbol{\pi} \ll \mathbf{1}$. 下面证明 $\boldsymbol{\pi}$ 比其他正的固定点都来得小. 设 $\boldsymbol{\pi}^* > 0$ 满足 $\boldsymbol{\phi}(\boldsymbol{\pi}^*) = \boldsymbol{\pi}^*$. 由单调性知, 我们有 $\boldsymbol{\pi}^* = \boldsymbol{\phi}(\boldsymbol{\pi}^*) \geqslant \boldsymbol{\phi}_1(0,0)$, 由迭代得 $\boldsymbol{\pi}^* \geqslant \boldsymbol{\phi}_n(0,0)$, 取极限后即得到所希望的结果: $\boldsymbol{\pi}^* \geqslant \boldsymbol{\pi}$. ■ 410

我们可以进一步加强定理 5.1 的结果.

定理 5.2 在定理 5.1 的假设下, 如果 $\boldsymbol{q}$ 是单位正方形上除 **1** 之外任意向量, 则 $\lim_{n\to\infty}\boldsymbol{\phi}_n(\boldsymbol{q})=\boldsymbol{\pi}$.

证明 首先假设 $0\leqslant q^i<1(i=1,2)$. 若 N 是正整数, 则 $\varphi_n^{(1)}(\boldsymbol{q})$ 的泰勒展开式可表为如下形式:

$$\begin{aligned}\varphi_n^{(1)}(\boldsymbol{q})=&\Pr\{|\boldsymbol{X}_n|=0|U_0=1,V_0=0\}\\&+\sum_{0<|X|\leqslant N}\Pr\{\boldsymbol{X}_n=\boldsymbol{x}|U_0=1,V_0=0\}(q^1)^{x_1}(q^2)^{x_2}\\&+\sum_{|X|>N}\Pr\{\boldsymbol{X}_n=\boldsymbol{x}|U_0=1,V_0=0\}(q^1)^{x_1}(q^2)^{x_2}.\end{aligned}$$

上面后一个和式有界于 $(\max(q^1,q^2))^N\Pr\{|\boldsymbol{X}_n|>N\}\leqslant(\max(q^1,q^2))^N$, 由于 $\max(q^1,q^2)<1$, 因而当 $N\to\infty$ 时, 此界趋于 0.

因为任何有限非零状态都是瞬态, $|\boldsymbol{X}_n|$ 或趋于 0 或趋于 ∞. 所以, 当 $n\to\infty$ 时第一个和式的每个系数均趋于 0. 由此推得当 $n\to\infty$ 且 N 固定时, 第一个和式趋于 0. 这就证明了

$$\begin{aligned}\lim_{n\to\infty}\varphi_n^{(1)}(\boldsymbol{q})&=\lim_{n\to\infty}\Pr\{|\boldsymbol{X}_n|=0|U_0=1,V_0=0\}\\&=\lim_{n\to\infty}\varphi_n^{(1)}(\boldsymbol{0})=\pi^{(1)}.\end{aligned}$$

类似可证

$$\lim_{n\to\infty}\varphi_n^{(2)}(\boldsymbol{q})=\pi^{(2)}.$$

若 $q^{(i)}(i=1,2)$ 有且仅有一个等于 1, 则 $\boldsymbol{\phi}_1(\boldsymbol{q})=(\varphi^1(\boldsymbol{q}),\varphi^2(\boldsymbol{q}))$ 确定每个分量均为严格小于 1 的非负向量. 我们对 $\boldsymbol{\phi}_1(\boldsymbol{q})$ 使用上述的分析方法, 得

$$\lim_{n\to\infty}\boldsymbol{\phi}_n(\boldsymbol{\phi}_1(\boldsymbol{q}))=\boldsymbol{\pi}=\lim_{n\to\infty}\boldsymbol{\phi}_{n+1}(\boldsymbol{q}).$$ ■

推论 5.1 方程 (5.5) 的非负解必是 **1** 和 $\boldsymbol{\pi}$.

8.6 多维分支过程

在细节上作些修改, 可把前一节的结果推广到多维分支过程. 它们的证明过程 411 并无新鲜东西. 以下我们仅列出结果, 勤奋的读者可补出详细证明.

我们考虑由 p 个类型构成的分支过程, 不同的类型对应于生物实际不同的变异形式, 也可看作单一生物的不同年龄或其他某个性质. 限制类型数为有限值的要求, 不妨可解释为我们把年龄分为有限多个年龄段.

在宇宙射线的电子 — 光子级联中, 不同类型可以解释为光子的不同能量等级. 与类型 i 相联系的是概率母函数

$$f^{(i)}(s_1,\cdots,s_p)=\sum_{r_1,\cdots,r_p=0}^{\infty}p^{(i)}(r_1,\cdots,r_p)s_1^{r_1},\cdots,s_p^{r_p},$$
$$|s_1|\leqslant 1,\cdots,|s_p|\leqslant 1\quad i=1,\cdots,p.$$

其中 $p^{(i)}(r_1,\cdots,r_p)$ 表示一个 i 型单个体产生 r_1 个 1 型, r_2 个 2 型, $\cdots r_p$ 个 p 型后代的概率.

引入向量记号: $\boldsymbol{s}=(s_1,\cdots,s_p)$.

设 $f_n^{(i)}(\boldsymbol{s})$ 表示由一个 i 型个体产生第 n 代的概率母函数. 类似于 (5.3) 我们有

$$f_{n+1}^{(i)}(\boldsymbol{s})=f^{(i)}(f_n^{(1)}(\boldsymbol{s}),f_n^{(2)}(\boldsymbol{s}),\cdots,f_n^{(p)}(\boldsymbol{s})),\quad f_0^{(i)}(\boldsymbol{s})=s_i,n=0,1,\cdots,i=1,\cdots,p,$$

设 $\boldsymbol{Z}_n=(Z_n^{(1)},\cdots,Z_n^{(p)})$ 表示第 n 代群体各型 (共 p 个型) 个体含量的向量. 类似于 (5.4) 有

$$E[\boldsymbol{Z}_{n+m}|\boldsymbol{Z}_n]=\boldsymbol{Z}_n\boldsymbol{M}^m,$$

其中 $\boldsymbol{M}=\|m_{ij}\|_{i,j=1}^p$ 是一阶矩矩阵:

$$m_{ij}=E[\boldsymbol{Z}_1^{(j)}|\boldsymbol{Z}_0=\boldsymbol{e}_i]=\frac{\partial f^{(i)}}{\partial s_j}(1,1,\cdots,1),i,j=1,\cdots,p.$$

此处 $\boldsymbol{e}_i$ 表示第 i 个分量为 1 其余分量为 0 的向量.

下来我们来叙述对于 p 维分支过程类似于定理 5.1 的结果. 假设对所有 $i,j,m_{ij}>0$. (只要对某个 n 和所有 i,j 有 $m_{ij}^{(n)}>0$ 即足够.) 设 $\pi^{(i)}$ 表示初始时刻只有一个 i 型. $(i=1,\cdots,p)$ 个体时的消失概率, 即

$$\pi^{(i)}=\Pr\{\boldsymbol{Z}_n=0\text{ 对某个 }n|\boldsymbol{Z}_0=e_i\}.$$

用 $\boldsymbol{\pi}$ 表示向量 $(\pi^{(1)},\pi^{(2)},\cdots,\pi^{(p)})$, 用 $\boldsymbol{1}$ 表示向量 $(1,1,\cdots,1)$.

定理 6.1 设对所有 $i,j=1,\cdots,p$ 有 $m_{ij}>0,\rho$ 表示矩阵 $\boldsymbol{M}$ 的绝对值最大的特征值. 若 $\rho\leqslant 1$, 则 $\boldsymbol{\pi}=\boldsymbol{1}$; 若 $\rho>1$, 则 $\boldsymbol{0}\leqslant\boldsymbol{\pi}\ll\boldsymbol{1}$ 且 $\boldsymbol{\pi}$ 满足方程

$$\pi^{(i)}=f^{(i)}(\boldsymbol{\pi}),\quad i=1,\cdots,p.$$

8.7 连续时间分支过程

8.1~8.6 节中所研究的分支过程限于发生的时间是固定的情况. 虽然某些现象, 特别是实验室试验, 是符合这种情况的, 但大多数自然界的繁殖过程关于时间都是 412

连续发生的. 因此我们有必要研究连续时间的分支过程.

本节将考虑时间齐次马尔可夫分支过程的构造. 在 8.11 节中马尔可夫的限制将被去掉. 我们用指定过程的无限小概率的办法来确定状态变量为 $X(t)=\{$在时刻 t 粒子的数目, 已知 $X(0)=1\}$的连续时间马尔可夫分支过程. 令

$$\delta_{1k}+a_k h+o(h), \quad k=0,1,2,\cdots, \tag{7.1}$$

(见 4.4 节和第 2 卷第 14 章) 表示在很小时间区间 $(t,t+h)$ 内, 单个粒子分裂产生 k 个粒子 (或研究对象) 的概率. 在式 (7.1) 中, δ_{1k} 照例为克罗内克 δ 符号, 并且假设 $a_1\leqslant 0, a_k\geqslant 0$, $k=0,2,3,\cdots$, 及

$$\sum_{k=0}^{\infty} a_k=0 \tag{7.2}$$

我们进一步假定每个粒子彼此相互独立地服从由式 (7.1) 给出的无限小概率分布. 注意, a_k 不依赖于分裂或转换的时间, 因此实际上我们可假定转移概率关于时间是齐次的.

表示无限小转移的另一途径是区分分裂发生的时间和分裂的情况. 这样, 每个对象的生活时间是随机变量, 它服从以均值为 $\lambda^{-1}=a_0+a_2+a_3+\cdots$ 的指数分布, 当其生命结束时, 它产生数目为 D 的同类对象的后代, 其中 D 为随机变量, 其分布概率为

$$\Pr\{D=k\}=\frac{a_k}{a_0+a_2+a_3+\cdots}, \quad k=0,2,3,\cdots.$$

各个个体的寿命和后代分布均是独立同分布的. 考虑到独立性假设, 特别是各个个体行为的独立性, 我们可以把式 (7.1) 等价地用无限小转移概率矩阵表示为

$$\Pr\{X(t+h)=n+k-1|X(t)=n\}=na_k h+o(h) \tag{7.3}$$

并且

$$\Pr\{X(t+h)=n|X(t)=n\}=1+na_1 h+o(h), \tag{7.4}$$

413 其中, $o(h)/h$ 当 $h\to 0_+$ 时趋向于 0.

我们早已在生灭过程那里遇见连续时间分支过程的例子. 事实上, 若置 $a_2=\lambda, a_0=\mu, a_1=-(\lambda+\mu)$ 和 $a_k=0(k=3,4,\cdots)$, 则 $(\lambda+\mu)^{-1}$ 可以解释为一个生或灭事件的概率; $\lambda/(\lambda+\mu)(\mu/(\lambda+\mu))$ 是在一个生或灭的事件已发生条件下生 (灭) 的概率. 这样得到的随机过程, 其状态变量是群体含量, 现可看为是线性增长的生灭过程 (见 4.6 节).

在第 14 章我们将解释, 若构造一个马尔可夫过程, 使其对应于一事先指定的无限小概率矩阵, 将是件困难事情. 而要保证所构造过程服从分支过程规则, 即各个粒

子产生独立的家族, 并且其后代也是相互独立地繁殖, 等等, 是件更加困难的事情. 我们将不进行这方面的分析, 因为这已超出本书的范围. 读者若要了解这一点, 可参阅 Harris 的书 (见本章末参考书目). 此外, 我们提醒读者注意第 2 卷第 14 章中有关马尔可夫过程和无限小概率矩阵之间关系的讨论.

设 $p_{ij}(t)$, 以后也将如此表示, 表示在时刻 o 群体有 i 个个体, 而在时刻 t 群体有 j 个个体的概率, 用符号表示为 $p_{ij}(t) = \Pr\{X(t+s) = j|X(s) = i\}$. 如记号所示, 这个概率仅依赖于间隔时间, 即过程具有平稳转移概率. 我们引进母函数

$$\phi(t;s) = \sum_{j=0}^{\infty} P_{1j}(t)s^j. \tag{7.5}$$

由于个体的行是相互独立的, 我们有基本关系式

$$\sum_{j=0}^{\infty} P_{ij}(t)s^j = [\phi(t;s)]^i. \tag{7.6}$$

公式 (7.6) 刻划了分支过程并使其与别的连续时间马尔可夫过程区别开来. 它表示不同个体 (或粒子) 各自引起独立的过程实现的性质, 不受其他个体繁殖的后代影响. 换言之, i 个初始个体经过 t 时间, 繁殖成拥有含量为 $X(t;i)$ 的群体依概率意义与 i 个由单个个体经 t 时间的繁殖而成的群体的总和相同.

由于对时间的齐次性, 切普曼 - 科尔莫戈罗夫方程取如下形式:

$$P_{ij}(t+\tau) = \sum_{k=0}^{\infty} p_{ik}(t)P_{kj}(\tau) \tag{7.7}$$

414

由 (7.5), (7.6) 和 (7.7) 得

$$\begin{aligned}
[\phi(t+\tau;s)]^i &= \sum_{j=0}^{\infty} P_{ij}(t+\tau)s^j = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} P_{ik}(t)P_{kj}(\tau)s^j. \\
&= \sum_{k=0}^{\infty} P_{ik}(t)\sum_{j=0}^{\infty} P_{kj}(\tau)s^j = \sum_{k=0}^{\infty} P_{ik}(t)[\phi(\tau;s)]^k \\
&= [\phi(t;\phi(\tau;s))]^i,
\end{aligned}$$

特别有

$$\phi(t+\tau;s) = \phi(t;\phi(\tau;s)). \tag{7.8}$$

连续时间情形下的关系式 (7.8) 类似于离散时间分支过程情形下的迭代公式 (见 8.2 节) 下面我们介绍由 (7.1) 定义的无限小概率的母函数. 令

$$u(s) = \sum_{k=0}^{\infty} a_k s^k.$$

考虑

$$\phi(h;s)=\sum_{j=0}^{\infty}P_{1j}(h)s^j=\sum_{j=0}^{\infty}(\delta_{1j}+a_jh+o(h))s^j \tag{7.9}$$

$$=s+h\sum_{j=0}^{\infty}a_js^j+o(h)=s+hu(s)+o(h).$$

在 (7.8) 中, 令 $\tau=h$ 得

$$\phi(t+h;s)=\phi(t;\phi(h;s))=\phi(t;s+hu(s)+o(h)),$$

依泰勒定理, 将上式右边关于第二个变量展开得

$$\phi(t+h;s)=\phi(t;s)+\frac{\partial\phi(t;s)}{\partial s}hu(s)+o(h).$$

则

$$\frac{\phi(t+h;s)-\phi(t;s)}{h}=\frac{\partial\phi(t;s)}{\partial s}u(s)+\frac{o(h)}{h}.$$

令 h 递减趋于 0, 导出

$$\frac{\partial\phi(t;s)}{\partial t}=\frac{\partial\phi(t;s)}{\partial s}u(s). \tag{7.10}$$

这是一个关于二变量函数 $\phi(t;s)$ 的偏微分方程, 其初始条件为

415

$$\phi(0;s)\equiv\sum_{j=0}^{\infty}P_{1j}(0)s^j\equiv s. \tag{7.11}$$

当 $u(s)$ 已知时, 由 (7.10), (7.11) 可解得 $\phi(t;s)$.

偏微分方程 (7.10) 只不过是科尔莫戈罗夫向前微分方程的另一种形式, 它为转移概率函数的母函数所满足.

我们还可导出相应于科尔莫戈罗夫向后微分方程且而被 ϕ 所满足的第二个微分方程. 为此, 在 (7.8) 中代入 $t=h$, 得

$$\phi(h+\tau;s)=\phi(h;\phi(\tau;s)),$$

然后利用前面 (7.9) 的泰勒展开式, 这就推得

$$\phi(h+\tau;s)=\phi(\tau;s)+hu(\phi(\tau;s))+o(h).$$

改写上式为

$$\frac{\phi(\tau+h;s)-\phi(\tau;s)}{h}=u(\phi(\tau;s))+\frac{o(h)}{h}. \tag{7.12}$$

令 $h \to 0_+$, 并用 t 代替 τ, 我们得到

$$\frac{\partial \phi(t;s)}{\partial t} = u(\phi(t;s)). \tag{7.13}$$

这是一个常微分方程, 其初始条件仍为 (7.11). 稍后, 我们将说明如何正确解 (7.13).

8.8 连续时间分支过程的消失概率

我们先来计算 $X(t)$ 的均值. 为此, 关于 s 并改变左边微分的顺序, 推得

$$\frac{\partial}{\partial t}\frac{\partial \phi(t;s)}{\partial s} = \frac{\partial^2 \phi(t;s)}{\partial s^2}u(s) + \frac{\partial \phi(t;s)}{\partial s}u'(s). \tag{8.1}$$

置 $s=1$, 由于 $u(1)=0$[即条件 (7.2)], 我们有

$$\frac{\partial m(t)}{\partial t} = u'(1)m(t), \tag{8.2}$$

其中

$$m(t) = EX(t) = \left.\frac{\partial \phi(t;s)}{\partial s}\right|_{s=1}$$

416

(8.2) 的解是

$$m(t) = \exp[u'(1)t], \tag{8.3}$$

这是因为, 我们若假设 $X(0)=1$, 其初始条件便为 $m(0)=1$.

其次, 我们来处理消失概率问题. 在此假定 $a_0>0$, 否则不可能消失. 我们只需考虑初始时刻只有一个个体的情况. 事实上, 由 (7.6) 我们有

$$\sum_{j=0}^{\infty} P_{ij}(t)s^j = \left[\sum_{j=0}^{\infty} P_{1j}(t)s^j\right]^i$$

因此,

$$P_{i0}(t) = [P_{10}(t)]^i.$$

但 $P_{i0}(t)$ 是含量为 i 个个体的群体经过时间 t 之后消失的概率, 直观上易知 $P_{i0}(t)$ 关于 t 是不减的. 下面我们将利用 (7.8) 严格证明这一事实. 实际上,

$$\begin{aligned} P_{i0}(t+\tau) = [\phi(t+\tau;0)]^i &= [\phi(t;\phi(\tau,0))]^i \\ &\geqslant [\phi(t,0)]^i = P_{i0}(t), \end{aligned}$$

其中我们使用了 $\phi(t,s)$ 关于 s 具有非负系数的幂级数的事实, 因此它关于 s 是增函数.

消失概率可定义为源于单个个体的“家族”最终绝灭的概率, 即

$$q = \lim_{t\to\infty} P_{10}(t).$$

利用离散时间分支过程的理论 (见 8.3 节), 我们易于确定连续时间情况下的消失概率. 设 t_0 是任意固定的正数并考虑离散时间过程

$$X(0), X(t_0), X(2t_0), \cdots, X(nt_0), \cdots,$$

其中 $X(t)$ 是原来的连续分支过程在时刻 t 时群体含量, 该群体在时刻 $t=0$ 只有一个个体. 由于 $X(t)$ 是一马尔可夫过程, 因此离散时间过程 $Y_n = X(nt_0)$ 显然是一马尔可夫链, 并且它还是一离散时间分支过程. 事实上, 由 $X(t)$ 的概率函数的齐次性的假设及式 (7.6), 我们有

$$\begin{aligned}\sum_{k=0}^{\infty} \Pr\{Y_{n+1}=k|Y_n=i\}s^k &= E[s^{Y_{n+1}}|Y_n=i]\\ &= E[s^{X((n+1)t_0)}|X(nt_0)=i] = E[s^{X(t_0)}|X(0)=i]\\ &= [\phi(t_0;s)]^i = \{E[s^{X(t_0)}|X(0)=1]\}^i\\ &= \{E[s^{Y_1}|Y_0=1]\}^i.\end{aligned}$$

417

这就证明了 Y_n 构成一个分支过程. 对于这个过程, 单独一个个体后代数目的母函数是 $\phi(t_0;s)$. 因此, 如 8.3 节中所证, 过程 Y_n 的消失概率是下面方程的最小非负根:

$$\phi(t_0;s) = s, \tag{8.4}$$

但

$$\begin{aligned}\Pr\{Y_n=0 \text{ 对于某个 } n\} &= \lim_{n\to\infty} \Pr\{Y_n=0\}\\ &= \lim_{n\to\infty} \Pr\{X(nt_0)=0\}\\ &= \lim_{t\to\infty} \Pr\{X(t)=0\} = q.\end{aligned}$$

因此, 连续时间分支过程 $X(t)$ 的消失概率 q 是方程 (8.4) 的最小非负根, 其中 t_0 是任意正数.

由于 q 是方程 (8.4) 对于任意 t_0 的根, 我们希望也能从一个不依赖于时间 t 的方程中来计算 q. 这确实可以做到, 我们有下面定理:

定理 8.1 消失概率 q 是方程

$$u(s) = 0 \tag{8.5}$$

的最小非负根. 因此, $q = 1$ 当且仅当 $u'(1) \leqslant 0$. (记注, $u(s) = \sum_{k=0}^{\infty} a_k s^k = a_1 s + [a_0 + a_2 s^2 + \cdots] = a_1 s + g(s)$.)

证明 由于 q 对于任意 t_0 满足 (8.4), 根据方程 (7.12), 对任意 $h > 0$, 我们有

$$0 = u(q) + \frac{o(h)}{h}.$$

令 $h \to 0_+$, 我们得到 $u(q) = 0$.

因为 $u''(s) = \sum_{k=2}^{\infty} a_k k(k-1) s^{k-2} \geqslant 0$, 所以 $u(s)$ 在区间 [0, 1] 是凸的. 由于 $u(1) = 0$ 和 $u(0) = a_0 > 0$, $u(s)$ 在 (0, 1) 至多有一个零点. 根据是否成立 $u'(1) \leqslant 0$ 或 $u'(1) > 0$, 我们分别有如图 8-3 或图 8-4 所示的情况. 注意, $E(X(t_0)) = E(Y) > 1$ 当且仅当 $u'(1) > 0$. 这意味着对于离散时间分支过程 $X(nt_0), n = 0, 1, 2, \cdots$ ($t_0 > 0$ 固定), 消失概率小于 1, 因此对于过程 $X(t)$ 也有相同结论. 在这种情况下, 消失概率一定是 $u(s)$ 在 $[0, 1]$ 上最小零点. 类似可推知, 若 $u'(1) \leqslant 0$, q 必定等于 1. 无论哪一种情况, q 均是 (8.5) 最小非负根.

418

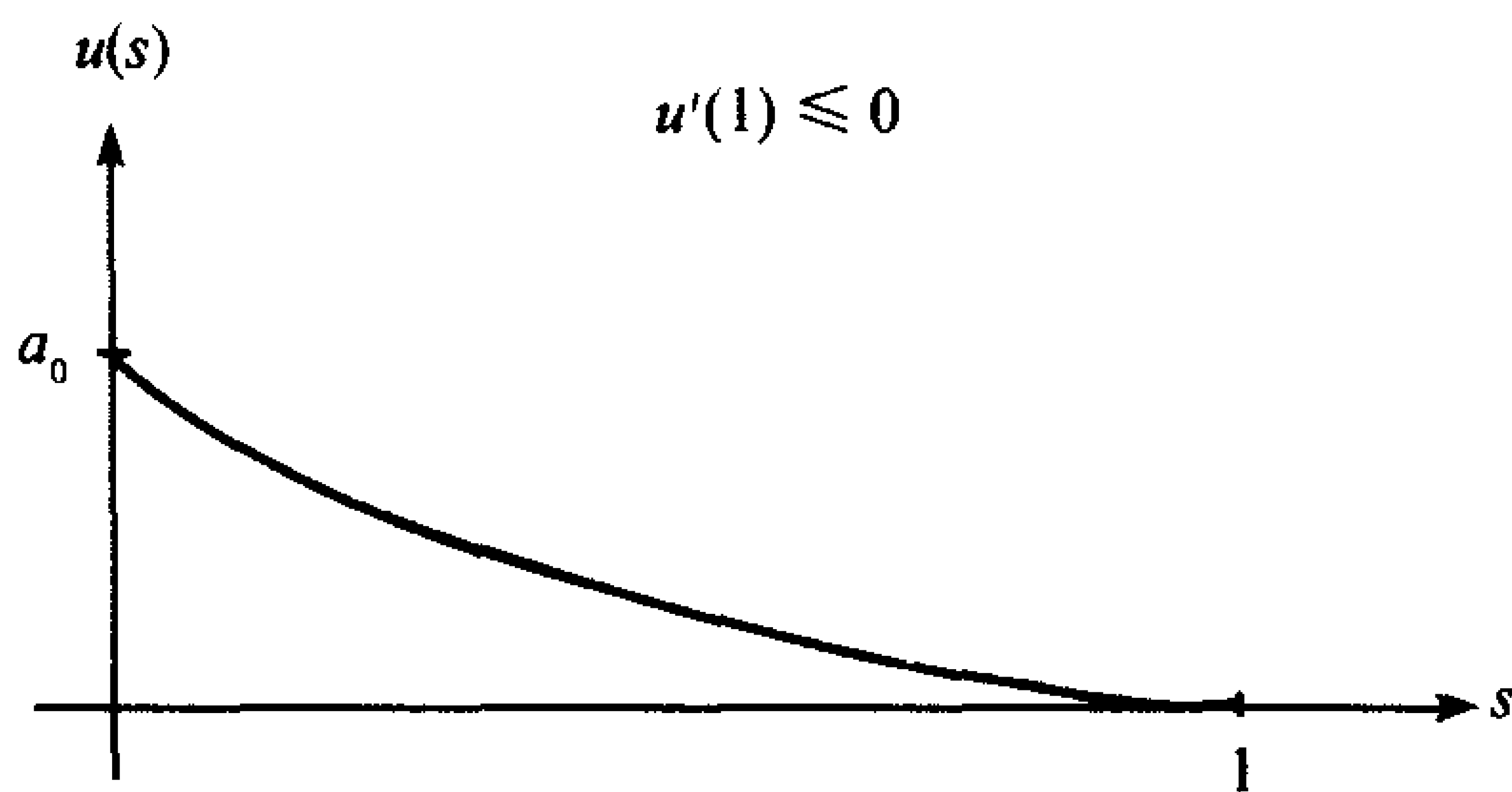


图 8-3

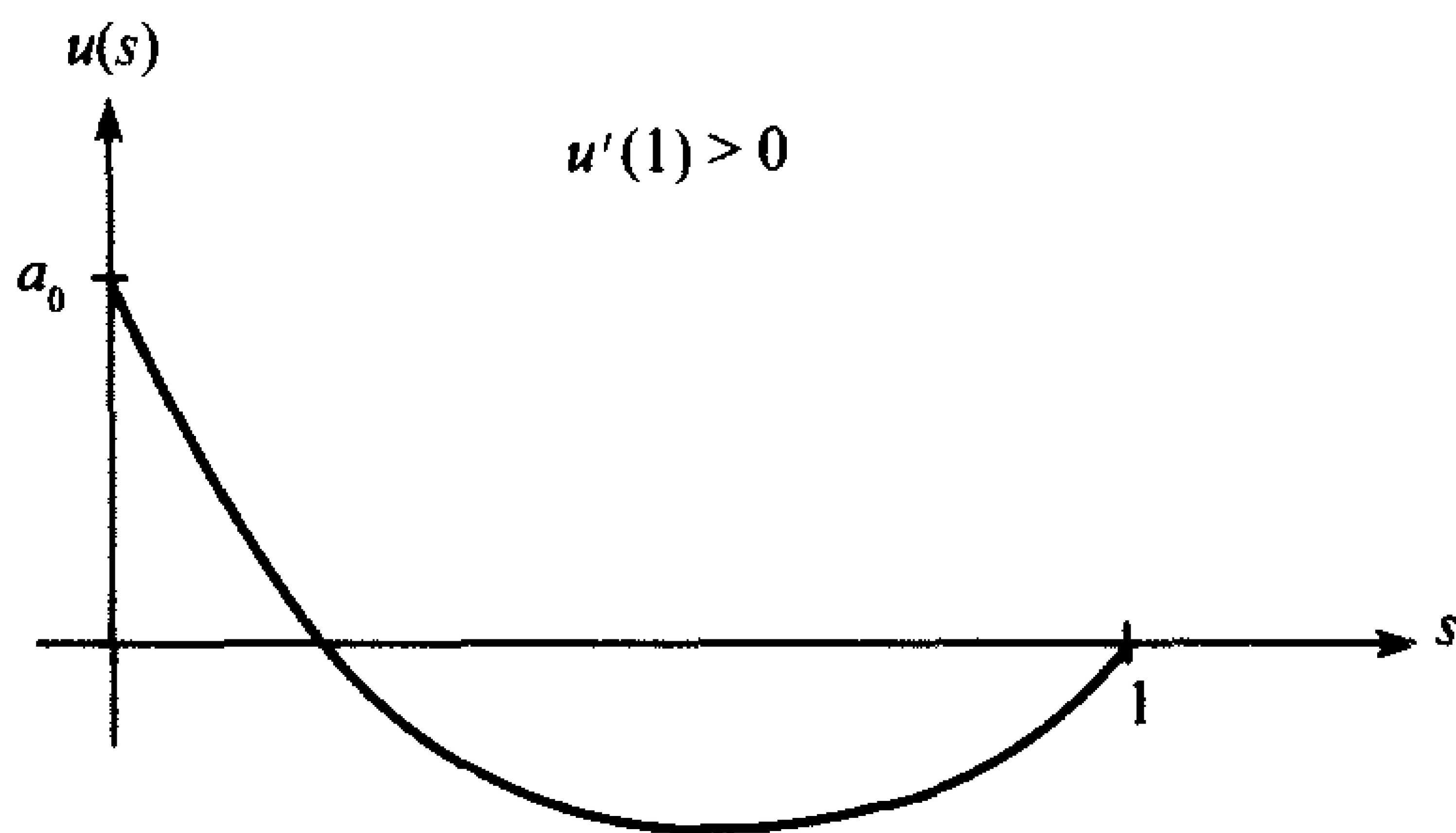


图 8-4

8.9 连续时间分支过程的极限定理

现在我们来解常微分方程 (7.13) 并分析当 $t \to \infty$ 时其增长情况. 鉴于 $\exp[u'(1)t]$ 是时刻 t 粒子数的期望值, 因此可推测知: 相应于 $u'(1)$ 是负的, 零或正的, $X(t)$ 具有不同的性态. 我们仅讨论 $u'(1) < 0$ 的情况并附加假设 $u''(1) < \infty$. 首先我们证明函数

$$B(s) = \frac{1}{u(s)} - \frac{1}{u'(1)(s-1)}$$

有界, 因此在 $0 \leqslant s < 1$ 可积. 事实上, 在 $s = 1$ 的邻域展开 $u(s)$, 得

$$u(s) = u(1) + u'(1)(s-1) + R(s)(s-1)^2, \quad s \leqslant 1,$$

其中,

419

$$\lim_{s \to 1^-} R(s) = \frac{u''(1)}{2!} < \infty. \tag{9.1}$$

注意到 $u(q) = u(1) = 0$, 我们有

$$\begin{aligned}\frac{1}{u(s)} &= \frac{1}{u'(1)(s-1) + R(s)(s-1)^2} = \frac{1}{u'(1)(s-1)} \cdot \frac{1}{1 + [R(s)(s-1)/u'(1)]} \\ &= \frac{1}{u'(1)(s-1)}\left\{1 - \frac{R(s)(s-1)/u'(1)}{1 + [R(s)(s-1)/u'(1)]}\right\}.\end{aligned}$$

因此,

$$B(s) = -\frac{R(s)/[u'(1)]^2}{1 + [R(s)(s-1)/u'(1)]} \tag{9.2}$$

由 (9.1) 直接可推得 $B(s)$ 是在 $s = 1-$ 的邻域内有界. 当然, 由 $B(s)$ 的定义知 $B(s)$ 在 $0 \leqslant s \leqslant 1 - \delta$ 有界, 这是因为 $u(s)$ 在所考虑情况 $[u'(1) < 0]$ 下仅当 $s = 1$ 时才为 0. 于是, 在条件 $u''(1) < \infty$ 和 $u'(1) < 0$ 之下, $B(s)$ 在 $0 \leqslant s < 1$ 是有界的. 现在我们对于 $0 \leqslant s < 1$ 可定义

$$K(s) = \int_1^s \left[\frac{1}{u(x)} - \frac{1}{u'(1)(x-1)}\right] \mathrm{d}x + \frac{\ln(1-s)}{u'(1)}, \tag{9.3}$$

此处积分存在且有限.

此外, 由假设 $u'(1) < 0$ 可得

$$K'(s) = \frac{1}{u(s)} > 0, \quad 0 \leqslant s < 1.$$

这表明 $K(s)$ 是严格递增的和连续的; 因此, 映射

$$w = K(s) \tag{9.4}$$

具有连续的、严格增加的反函数

$$s = K^{-1}(w) = L(w), \quad L(K(s)) = s, \tag{9.5}$$

且当 s 跑遍 $[0,1)$ 时, w 跑遍 $[k(0),\infty)$, 并有 $k(0)<0$. 现在我们已具备条件求初始条件 (7.11) 之下 (7.13) 的解. 对 (7.13) 分离变量并积分, 可导出关于 $\phi(t,s)$ 的隐式公式:

420

$$\int_s^{\phi(t;s)} \frac{\mathrm{d}x}{u(x)} = t.$$

利用 $k(\cdot)$ 定义, 我们得到

$$\begin{aligned}
t &= \int_s^{\phi(t;s)} \frac{\mathrm{d}x}{u(x)} = \int_s^{\phi(t;s)} \left[\frac{1}{u(x)} - \frac{1}{u'(1)(x-1)}\right]\mathrm{d}x + \frac{1}{u'(1)}\ln(1-x)\Big|_s^{\phi(t;s)} \\
&= \int_1^{\phi(t;s)} \left[\frac{1}{u(x)} - \frac{1}{u'(1)(x-1)}\right]\mathrm{d}x + \frac{\ln(1-\phi(t;s))}{u'(1)} \\
&\quad - \int_1^{s} \left[\frac{1}{u(x)} - \frac{1}{u'(1)(x-1)}\right]\mathrm{d}x - \frac{\ln(1-s)}{u'(1)} \\
&= K(\phi(t;s)) - K(s)
\end{aligned}$$

即

$$K(\phi(t;s)) = t + K(s).$$

由于 $K(\cdot)$ 存在反函数, 因此对于 $0 \leqslant s < 1$ 和 $t \geqslant 0$ 有

$$\phi(t;s) = K^{-1}(t + K(s)). \tag{9.6}$$

在假设 $u'(1)<0, u''(1)<\infty$ 之下, 我们还可以导出消失概率关于时间 t 的某些渐近结果. 因为 $B(s)$ 在 $0 \leqslant s < 1$ 是有界的, 并且 $\lim\limits_{s\to 1-} B(s)$ 存在, 在 $s=1$ 的邻域内我们有 [见 (9.3)]

$$K(s) = \frac{\ln(1-s)}{u'(1)} - C\cdot(1-s) + o(1-s). \tag{9.7}$$

此处 C 是负常数; 事实上, $\lim\limits_{x\to 1-}[1/u(x) - (1/(x-1)u'(1))] = C = (-u''(1)/2[u'(1)]^2)$. $o(1-s)$ 按通常意义理解, 即 $[o(1-s)]/(1-s) \to 0$(当 $s\to 1-$ 时). 改写 (9.7) 为

$$\ln(1-s) = u'(1)K(s) - Cu'(1)(1-s) + o(1-s)$$

取指数得

$$1 - s = \exp[u'(1)K(s)]\exp[-Cu'(1)(1-s)]\exp[o(1-s)].$$

但

$$\exp[o(1-s)] = 1 + o(1-s)$$

及

$$\exp[-Cu'(1)(1-s)] = 1 - Cu'(1)(1-s) + o(1-s),$$

由此推得

421

$$1 - s = \exp[u'(1)k(s)][1 - Cu'(1)(1-s) + o(1-s)]. \tag{9.8}$$

因此

$$\lim_{s\to 1-} \frac{1-s}{\exp[u'(1)K(s)]} = 1.$$

借助于此极限, 可把 (9.8) 写为

$$1 - s = \{\exp[u'(1)K(s)]\}\{1 - Cu'(1)\exp[u'(1)K(s)] + o(\exp[u'(1)K(s)])\}.$$

用 $K^{-1}(w)$ 代替 s, 我们得到 [见 (9.5)]

$$1 - K^{-1}(w) = \exp[u'(1)w][1 - Cu'(1)\exp[u'(1)w] + o(\exp[u'(1)w])], \tag{9.9}$$

此处 $s \to 1-$ 等价于 $w \to \infty$. 依 (9.6) 和 (9.9), 我们能够计算在时间 t 内不消失的概率. 即是

$$\begin{aligned}
1 - P_{10}(t) &= 1 - \phi(t;0) = 1 - K^{-1}(t + K(0)) \\
&= [\exp\{u'(1)(t + K(0))\}] \\
&\quad \times \{1 - Cu'(1)(\exp[u'(1)(t + K(0))]) + o(\exp[u'(1)(t + K(0))])\}] \\
&= \exp\{u'(1)K(0)\}\exp\{u'(1)t\} + o(\exp\{2u'(1)t\}) + o(\exp[u'(1)t]),
\end{aligned}$$

或等价地,

$$1 - P_{10}(t) = \exp[u'(1)K(0)]m(t) + o(\exp[u'(1)t]). \tag{9.10}$$

当 $t \to \infty$ 时另一渐近结果可用如下办法得到. 已知 $X(t) \neq 0$, $X(t)$ 的条件概率母函数定义为

$$g(z,t) = \sum_{k=0}^{\infty} \Pr\{X(t) = k | X(t) \neq 0\} z^k, \quad 0 \leqslant z < 1.$$

但

$$\begin{aligned}
\Pr\{X(t) = k | X(t) \neq 0\} &= \frac{\Pr\{X(t) = k, X(t) \neq 0\}}{\Pr\{X(t) \neq 0\}} \\
&= \begin{cases} 0 & \text{若 } k = 0, \\ \dfrac{\Pr\{X(t) = k\}}{1 - \Pr\{X(t) = 0\}} & \text{若 } k \neq 0. \end{cases}
\end{aligned}$$

据此

$$g(z;t)=\sum_{k=1}^{\infty}\frac{\Pr\{X(t)=k\}}{1-\Pr\{X(t)=0\}}z^k=\frac{\phi(t;z)-\phi(t;0)}{1-\phi(t;0)}$$
$$=\frac{K^{-1}(t+K(z))-K^{-1}(t+K(0))}{1-K^{-1}(t+K(0))}$$
$$=\frac{[1-K^{-1}(t+K(0))]-[1-K^{-1}(t+K(z))]}{1-K^{-1}(t+K(0))}$$

其中对 $\phi(t;s)$ 我们利用了式 (9.6).

422

由 (9.9) 得到 $K^{-1}(w)$ 的表达式代入上式, 我们有

$$g(z;t)=\frac{\mathrm{e}^{u'(1)(t+K(0))}[1+o(\mathrm{e}^{u'(1)(t+K(0))})]-\mathrm{e}^{u'(1)(t+K(z))}[1+o(\mathrm{e}^{u'(1)(t+K(z))})]}{\mathrm{e}^{u'(1)(t+K(0))}[1+o(\mathrm{e}^{u'(1)(t+K(0))})]}$$
$$=1-\mathrm{e}^{u'(1)[K(z)-K(0)]}\frac{1+o(\mathrm{e}^{u'(1)(t+k(z))})}{1+o(\mathrm{e}^{u'(1)(t+K(0))})}.$$

令 $t\to\infty$, 由于 $u'(1)<0$, 上式右边的分式趋于 1, 因此

$$\lim_{t\to\infty}g(z;t)=g(z)=1-\exp\{u'(1)[K(z)-K(0)]\}.$$

然而, 由 (9.3),

$$K(z)-K(0)=\int_0^z\left[\frac{1}{u(x)}-\frac{1}{u'(1)(x-1)}\right]\mathrm{d}x+\frac{\ln(1-z)}{u'(1)}$$
$$=\int_0^z\frac{\mathrm{d}x}{u(x)}-\frac{\ln(1-x)}{u'(1)}\bigg|_0^z+\frac{\ln(1-z)}{u'(1)}=\int_0^z\frac{\mathrm{d}x}{u(x)}.$$

因此, 当 $t\to\infty$ 时, 得到极限概率母函数

$$g(z)=1-\exp\left[u'(1)\int_0^z\frac{\mathrm{d}x}{u(x)}\right]=\sum_{k=1}^{\infty}\lim_{t\to\infty}\Pr\{X(t)=k|X(t)\neq 0\}z^k.$$

综上所述, 有

定理 9.1 设 $X(t)$ 是连续时间分支过程, 其无限小概率母函数为

$$u(s)=\sum_{k=0}^{\infty}a_k s^k, \tag{9.11}$$

其中 $\{a_k\}$ 具有 (7.1) 所述意义并满足条件 (7.2). 若 $u''(1)<\infty$, 并假设 $u(1)<0$, 使得其消失概率 $q=1$(见定理 8.1). 则

$$\phi(t,s)=\sum_{k=0}^{\infty}\Pr\{X(t)=k|X(0)=1\}s^k \tag{9.12}$$
$$=K^{-1}(t+K(s)),\quad t\geqslant 0,|s|<1,$$

其中 $K(s)$ 由 (9.3) 确定, 至时刻 t 的不消失概率依如下指数速度趋于 0

423

$$\lim_{t\to\infty}\frac{1-P_{10}(t)}{\exp[u'(1)K(0)]\exp[u'(1)t]}=1.$$

此外, 随机变量 $X(t)$ 在条件 $X(t)>1$ 下有极限分布, 其概率母函数为

$$\begin{aligned}g(z,t)&=\frac{\sum_{k=1}^{\infty}\Pr\{X(t)=k|X(0)=1\}z^k}{1-\Pr\{X(t)=0\}}\\&\to 1-\exp\left[u'(1)\int_0^z\frac{\mathrm{d}x}{u(x)}\right],\quad (\text{当 } t\to\infty \text{ 时}).\end{aligned}\tag{9.13}$$

我们将不加证明地叙述对应于情况 $u'(1)=0$ 和 $u'(1)>0$ 的极限定理. 它们的证明比较复杂, 虽然本质上是类似的.

定理 9.2 (i) 若 $u'(1)=0, u''(1)<\infty$, 则

$$\Pr\{X(t)>0|X(0)=1\}\sim\frac{2}{u''(t)}\frac{1}{t},\quad t\to\infty,$$

且

$$\lim_{t\to\infty}\Pr\left\{\frac{2X(t)}{u''(1)t}>\lambda\middle|X(t)>0\right\}=\mathrm{e}^{-\lambda},\quad \lambda>0.$$

(ii) 若 $u'(1)>0$ 和 $u''(1)<\infty$, 则

$$Z(t)=\frac{X(t)}{\exp[u'(1)t]}$$

当 $t\to\infty$ 时, 具有极限分布.

8.10 二维连续时间分支过程

考虑两种不同类型的粒子, 我们分别称其为 1 型和 2 型粒子. 关于两种类型粒子的连续时间分支马尔可夫过程将由适当指定的无限小概率 [(7.3) 和 (7.4)] 所确定. 详细地说, 我们假定每个 $i(i=1,2)$ 型的粒子在任一小区间 $(t,t+h)$ 转换为 k_1 个 1 型的粒子和 k_2 个 2 型的粒子 $(k_1,k_2,=0,1,2,\cdots)$ 的概率分别是

$$\delta_{1k_1}\delta_{0k_2}+a^{(1)}_{k_1,k_2}h+o(h)\quad (\text{对于 1 型})$$

和

424

$$\delta_{0k_1}\delta_{1k_2}+a^{(2)}_{k_1,k_2}h+o(h)\quad (\text{对于 2 型})$$

其中 δ_{ij} 是熟知的克罗内克 δ 符号. 我们假设粒子在任何时刻的转换都与它的过去独立, 也独立于其他粒子 (不管是哪一型) 的过去和现在. 注意, 这里我们再次假定了转移概率关于时间是齐次的, 因此参数 $a_{k_1,k_2}^{(i)}$ 与时间无关. 这些参数必须满足条件:

$$a_{1,0}^{(1)} \leqslant 0, \quad a_{01}^{(2)} \leqslant 0,$$
$$a_{k_1,k_2}^{(1)} \geqslant 0 \quad \text{对所有 } k_1,k_2=0,1,2,\cdots, \text{但 } k_1=1,k_2=0 \text{ 除外}.$$
$$a_{k_1,k_2}^{(2)} \geqslant 0 \quad \text{对所有 } k_1,k_2=0,1,2,\cdots. \text{但 } k_1=0,k_2=1 \text{ 除外}.$$

和

$$\sum_{k_1,k_2=0}^{\infty} a_{k_1,k_2}^{(i)} = 0, \quad i=1,2.$$

我们引入两个无限小母函数

$$u^{(i)}(s_1,s_2) = \sum_{k_1,k_2=0}^{\infty} a_{k_1,k_2}^{(i)} s_1^{k_1} s_2^{k_2}, \quad i=1,2,(|s_1| \leqslant 1, |s_2| \leqslant 1).$$

设 $P_{k_1,k_2,;j_1,j_2}(t)$ 表示在时刻 0 有 k_1 个 1 型和 k_2 个 2 型的粒子的群体经过时间 t 后转换为 j_1 个 1 型和 j_2 个 2 型的粒子的概率. 由于无限小概率 $a_{k_1,k_2}^{(i)}$ 是与时间无关的, 故转移概率必关于时间是齐次的. 我们定义概率母函数

$$\phi^{(1)}(t;s_1,s_2) = \sum_{j_1,j_2=0}^{\infty} P_{1,0;j_1,j_2}(t) s_1^{j_1} s_2^{j_2},$$

$$\phi^{(2)}(t;s_1,s_2) = \sum_{j_1,j_2=0}^{\infty} P_{0,1;j_1,j_2}(t) s_1^{j_1} s_2^{j_2}.$$

与一维分支过程类似, 可推得

$$\sum_{j_1,j_2=0}^{\infty} P_{k_1,k_2;j_1,j_2}(t) s_1^{j_1} s_2^{j_2} = [\phi^{(1)}(t;s_1,s_2)]^{k_1} [\phi^{(2)}(t;s_1,s_2)]^{k_2}, \quad (k_1,k_2,=0,1,2,\cdots). \tag{10.1}$$

事实上, 关系式 (10.1) 可以作为二维连续时间分支过程的定义. 换言之, 任何转移概率矩阵若满足 (10.1) 就说它产生一个二维连续时间马尔可夫分支过程. 过程的马尔可夫特征体现于切普曼 - 科尔莫戈罗夫方程式

$$P_{k_1,k_2;j_1,j_2}(t+\tau) = \sum_{l_1,l_2=0}^{\infty} P_{k_1,k_2;l_1,l_2}(t) P_{l_1,l_2;j_1,j_2}(\tau). \tag{10.2}$$

由 (10.1) 和 (10.2) 我们有

$$\phi^{(1)}(t+\tau;s_1,s_2) = \sum_{j_1,j_2=0}^{\infty} P_{1,0;j_1,j_2}(t+\tau) s_1^{j_1} s_2^{j_2}$$

$$
\begin{aligned}
&= \sum_{j_1,j_2=0}^{\infty} \sum_{l_1,l_2=0}^{\infty} P_{1,0;l_1,l_2}(t) P_{l_1,l_2;j_1,j_2}(\tau) s_1^{j_1} s_2^{j_2} \\
&= \sum_{l_1,l_2=0}^{\infty} P_{1,0;l_1,l_2}(t) \sum_{j_1,j_2=0}^{\infty} P_{l_1,l_2;j_1,j_2}(\tau) s_1^{j_1} s_2^{j_2} \\
&= \sum_{l_1,l_2=0}^{\infty} P_{1,0;l_1,l_2}(t) [\phi^{(1)}(\tau; s_1, s_2)]^{l_1} [\phi^{(2)}(\tau; s_1, s_2)]^{l_2} \\
&= \phi^{(1)}(t; \phi^{(1)}(\tau; s_1, s_2), \phi^{(2)}(\tau; s_1, s_2)).
\end{aligned}
$$

对于母函数 $\phi^{(2)}(t; s_1, s_2)$ 同理可得

$$
\phi^{(i)}(t+\tau; s_1, s_2) = \phi^{(i)}(t; \phi^{(1)}(\tau; s_1, s_2), \phi^{(2)}(\tau; s_1, s_2)), (i = 1, 2). \tag{10.3}
$$

此外,

$$
\begin{aligned}
\phi^{(1)}(h; s_1, s_2) &= \sum_{j_1,j_2=0}^{\infty} P_{1,0;j_1,j_2}(h) s_1^{j_1} s_2^{j_2} \\
&= \sum_{j_1,j_2=0}^{\infty} [\delta_{1j_1} \delta_{0,j_2} + a_{j_1 j_2}^{(1)} h + o(h)] s_1^{j_1} s_2^{j_2} \\
&= s_1 + h u^{(1)}(s_1, s_2) + o(h),
\end{aligned}
$$

对 $\phi^{(2)}(h; s_1, s_2)$ 也有类似式子, 因此

$$
\phi^{(i)}(h; s_1, s_2) = s_i + h u^{(i)}(s_1, s_2) + o(h), \quad i = 1, 2. \tag{10.4}
$$

现在我们来推导 $\phi^{(i)}(t; s_1, s_2)$ 所满足的类似于 (7.10) 和 (7.13) 的偏微分方程. 为此, 置 $\tau = h$, 并把 (10.4) 代入 (10.3). 应用泰勒展开得

$$
\begin{aligned}
\phi^{(i)}(t+h; s_1, s_2) &= \phi^{(i)}(t; s_1 + h u^{(1)}(s_1, s_2) + o(h), s_2 + h u^{(2)}(s_1, s_2) + o(h)) \\
&= \phi^{(i)}(t; s_1, s_2) + \frac{\partial \phi^{(i)}(t; s_1, s_2)}{\partial s_1} h u^{(1)}(s_1, s_2) \\
&\quad + \frac{\partial \phi^{(i)}(t; s_1, s_2)}{\partial s_2} h u^{(2)}(s_1, s_2) + o(h).
\end{aligned}
$$

426

两边除以 h, 并令 $h \to 0$, 我们形式上得到微分方程

$$
\begin{aligned}
\frac{\partial \phi^{(i)}(t; s_1, s_2)}{\partial t} = &\frac{\partial^{(i)}(t; s_1, s_2)}{\partial s_1} u^{(1)}(s_1, s_2) \\
&+ \frac{\partial \phi^{(i)}(t; s_1, s_2)}{\partial s_2} u^{(2)}(s_1, s_2), i = 1, 2.
\end{aligned} \tag{10.5}
$$

我们再次从 (10.3) 开始, 此时令 $t=h$ 并利用 (10.4), 推得

$$\begin{aligned}\phi^{(i)}(h+\tau;s_1,s_2)&=\phi^{(i)}(h;\phi^{(1)}(\tau;s_1,s_2),\phi^{(2)}(\tau;s_1,s_2))\\&=\phi^{(i)}(\tau;s_1,s_2)+hu^{(i)}(\phi^{(1)}(\tau;s_1,s_2),\phi^{(2)}(\tau;s_1,s_2))+o(h).\end{aligned}$$

两边除以 h, 令 $h\to o$, 并用 t 代替 τ, 我们得到第二个微分方程组:

$$\frac{\partial\phi^{(i)}(t;s_1,s_2)}{\partial t}=u^{(i)}(\phi^{(1)}(t;s_1,s_2),\phi^{(2)}(t;s_1,s_2)),\quad i=1,2.\tag{10.6}$$

对于 (10.5) 和 (10.6) 的初始条件都是

$$\phi^{(i)}(0;s_1,s_2)=s_i,\quad i=1,2.$$

利用 (10.5) 和 (10.6) 可推导出该过程随机变量的矩所满足的常微分方程组, 对此我们不加详述.

下面我们介绍一些二维连续时间分支过程的例子和应用.

例 1 第一个例子涉及有迁入的分支过程. 考虑一维连续时间分支过程, 除分支外同时允许某些粒子迁入该系统. 回忆前面的记号

$$\delta_{1k}+a_kh+o(h),\quad k=0,1,2,\cdots,$$

表示一个粒子在较小时间区间 $(t,t+h)$ 内转换为 k 个粒子的概率, 此事件与该粒子的过去以及其他粒子都是独立的. 现在我们把迁入因素添加进去, 设

$$\delta_{0k}+b_kh+o(h),\quad k=0,1,2,\cdots,$$

表示在时间区间 $(t,t+h)$ 内, 独立于群体的现在与过去的 k 个同类型粒子迁入且与该群体合并在一起的概率, 注意, 参数 a_k 和 b_k 均假定独立于转换及迁入发生的时间, 即联系每个个体的无限小转移概率是时间齐次的. a_k 和 b_k 必须满足条件:

427

$$\begin{aligned}&a_1\leqslant 0,\quad b_0\leqslant 0,\\&a_k\geqslant 0,\quad k=0,2,3,\cdots,\\&b_k\geqslant 0\quad k=1,2,3,\cdots,\\&\sum_{k=0}^{\infty}a_k=\sum_{k=0}^{\infty}b_k=0.\end{aligned}$$

令

$$\begin{aligned}P_k(t)&=\Pr\{\text{群体在时刻 } t=0 \text{ 无粒子存在, 而在时刻 } t \text{ 包含 } k \text{ 个粒子}\}\\&=\Pr\{X(t)=k|X(0)=0\},\quad k=0,1,2,\cdots,\end{aligned}\tag{10.7}$$

其母函数为

$$\phi(t;s)=\sum_{k=0}^{\infty}P_k(t)s^k. \tag{10.8}$$

我们目标是求 $P_k(t)$ 的值，若行不通，就确定它的某些性质。

引进两个无限小母函数

$$u(s)=\sum_{k=0}^{\infty}a_ks^k \quad 和 \quad v(s)=\sum_{k=0}^{\infty}b_ks^k.$$

我们可把带有迁入的一维连续时间分支过程视为二维连续时间分支过程作法如下：假定我们有两类质子：1 型和 2 型，它们具有特定的无限小转换概率，正如本章开始所述恒等式成立的理由如下：考虑两种类型的粒子：一种是实的，另一种是虚构的. 一个实的粒子至它的生命 (它的长短服从参数为 $\lambda^{-1}=a_0+a_2+a_3+\cdots$ 的指数分布) 结束时，以概率 $\lambda a_k(k=0,2,3,\cdots)$ 产生 k 个新的实粒子. 一个虚的粒子也有一个随机存活时间长度 (依参数为 $\bar{\lambda}^{-1}=b_1+b_2+\cdots$ 的指数分布)，生命结束时以概

428

率 $\bar{\lambda}\cdot b_l(l=1,2,3,\cdots)$ 产生 l 个实的粒子和一个虚的粒子. 注意，$\sum_{l=1}^{\infty}\bar{\lambda}b_l=1$. 虚构粒子的后代迁入因素. 置

$$a_{k_1,k_2}^{(1)}=\begin{cases}a_{k_1} & 若\ k_2=0,\\ 0 & 若\ k_2\neq 0,\end{cases}$$

$$a_{k_1,k_2}^{(2)}=\begin{cases}b_{k_1} & 若\ k_2=1,\\ 0 & 若\ k_2\neq 1.\end{cases}$$

依本节开始时记号，我们有

$$u^{(1)}(s_1,s_2)=\sum_{k_1,k_2=0}^{\infty}a_{k_1,k_2}^{(1)}s_1^{k_1}s_2^{k_2}=\sum_{k_1=0}^{\infty}a_{k_1}s_1^{k_1},$$

$$u^{(2)}(s_1,s_2)=\sum_{k_1,k_2=0}^{\infty}a_{k_1,k_2}^{(2)}s_1^{k_1}s_2^{k_2}=s_2\sum_{k_1=0}^{\infty}b_{k_1}s_1^{k_1}.$$

这样，

$$u^{(1)}(s_1,s_2)=u(s_1),\quad u^{(2)}(s_1,s_2)=s_2v(s_1).$$

在此特别情况下，微分方程 (10.5) 变为

$$\frac{\partial\phi^{(i)}(t;s_1,s_2)}{\partial t}=\frac{\partial\phi^{(i)}(t;s_1,s_2)}{\partial s_1}u(s_1)+\frac{\partial\phi^{(i)}(t;s_1,s_2)}{\partial s_2}s_2v(s_1),\quad i=1,2, \tag{10.9}$$

微分方程 (10.6) 变为

$$\frac{\partial\phi^{(1)}(t;s_1,s_2)}{\partial t}=u(\phi^{(1)}(t;s_1,s_2)) \tag{10.10}$$

和

$$\frac{\partial\phi^{(2)}(t;s_1,s_2)}{\partial t}=[\phi^{(2)}(t;s_1,s_2)]v(\phi^{(1)}(t;s_1,s_2)). \tag{10.11}$$

现在我们来说明二维分支过程的概率 $P_{0,1;j_1,j_2}(t)$ 与 (10.7) 所定义的概率之间的关系. 依照两种类型粒子的意义, 初始状态 (0, 1) 表示我们在时刻 0 没有实的粒子而存在由一个虚构粒子表示的潜在迁入因素. 由符号本身意义, 显然有

$$P_{0,1;j_1,j_2}(t)=\begin{cases} p_{j_1}(t) & \text{若 } j_2=1, \\ 0 & \text{若 } j_2\neq 1, \end{cases}$$

因此,

$$\phi^{(2)}(t;s_1,s_2)=s_2\phi(t;s_1). \tag{10.12}$$

429

由 (10.9), 用 s 代替 s_1, 我们得到

$$\frac{\partial\phi(t;s)}{\partial t}=\frac{\partial\phi(t;s)}{\partial s}u(s)+\phi(t;s)v(s). \tag{10.13}$$

初始条件为

$$\phi(0;s)=1. \tag{10.14}$$

在适当初始条件下解方程组 (10.10) 和 (10.11) 比解这个微分方程来得更容易. 方程 (10.10) 可以按分析 (7.13) 的方法加以处理. (10.10) 的解可表示为如 (9.6) 形式, 记其为 $f(t;s)$, 此处用 s 代替 s_1, 其中 s_2 不出现. 由于 (10.12) 的缘故, (10.11) 变为

$$\frac{\partial\phi(t;s)}{\partial t}=\phi(t;s)v(f(t;s)), \tag{10.15}$$

初始条件为 (10.14). 式 (10.15) 解为

$$\phi(t;s)=\exp\left[\int_0^t v(f(\tau;s))\mathrm{d}\tau\right].$$

例 2 本节末我们来描述一个简单的“二分裂”非马尔可夫的单型连续时间分支过程, 它可归结为二维连续时间马尔可夫分支过程. 假设一个粒子具有寿命分布密度

$$\frac{\lambda^2}{2}te^{-\lambda t}, \tag{10.16}$$

即是阶数为 2 的 Γ 分布. 当此粒子消失时, 它被两个相同类型粒子所代替, 每一个皆独立于另一个并与原来粒子相独立, 并且每一个都服从寿命分布 (10.16).

马尔可夫过程一般来说是用如下性质来刻划：在任何 e 知状态下其等候时间服从指数分布. 在本例中已知状态的等候时间由粒子寿命所确定. 如果它是指数分布则这些粒子的群体过程构成一马尔可夫过程. 但上述寿命分布不是指数分布, 而是两个指数分布的卷积.

430

设 $X(t)$ 表示时刻 t 的粒子数, 并假设 $X(0)=1$. 由于 (10.16) 是两个具有参数为 λ 的独立随机变量之和的密度函数, 我们可认为每个粒子的寿命经过两个分开的阶段, 在每个阶段均服从参数为 λ 的指数分布, 这样就很容易把它归结为二维连续时间马尔可夫分支过程. 此时我们不说是相同粒子的寿命经过两个阶段, 而说存在两种不同类型的粒子. 一个 1 型粒子具有参数为 λ 的指数分布寿命, 然后转换为 2 型粒子. 一个 2 型粒子具有参数为 λ 的指数分布寿命, 然后转换为两个 1 型的粒子, 这样, 利用本节开始时记号, 我们有

$$a_{k_1,k_2}^{(1)}=\begin{cases}-1 & \text{若 } k_1=1 \text{ 和 } k_2=0,\\ +1 & \text{若 } k_1=0 \text{ 和 } k_2=1,\\ 0 & \text{其他},\end{cases}$$

$$a_{k_1,k_2}^{(2)}=\begin{cases}+1 & \text{若 } k_1=2 \text{ 和 } k_2=0,\\ -1 & \text{若 } k_1=0 \text{ 和 } k_2=1,\\ 0 & \text{其他},\end{cases}$$

此处, 为简单起见, 假定 $\lambda=1$. 那么

$$u^{(1)}(s_1,s_2)=-s_1+s_2$$

且

$$u^{(2)}(s_1,s_2)=s_1^2-s_2.$$

方程式 (10.5) 和 (10.6) 变为特殊形式, 已知 $X(0)=1, X(t)$ 的母函数可由 $\phi^{(1)}(t;s_1,s_2)$ 置 $s_1=s_2=s$ 得到.

8.11 一般寿命的分支过程

本节所讨论的分支过程的每个对象 (或粒子、个体) 的寿命是服从一般分布的, 当其生命告终时产生它的后代. 这可与前后所介绍的寿命固定的或服从指数分布的分支过程相比较. 我们假定一个个体寿命随机变量 T 具有概率密度 $f(t)$, 即该个体在时间区间 $(t,t+\mathrm{d}t)$ 内死亡的概率为 $f(t)\mathrm{d}t$. 我们还假定该个体死亡时分裂成

两个, 从而产生了两个新的同类型个体, 它们的寿命分布相互独立, 而概率密度同样为 $f(t)$. 在它们生命结束时, 每一个又分裂成同类型的两个新个体, 如此连续以至无穷. 设 $N(t)$ 表示在时刻 t 存活个体的数目, 其概率分布记为 431

$$P_k(t) = \Pr\{N(t) = k\}, \quad k = 0, 1, 2, \cdots.$$

显然, 对所有 $t \geqslant 0$ 有 $P_0(t) = 0$, 因为群体在第一次分裂前恰好有一个个体, 第一次分裂后至少有两个个体, 从而在任何时刻至少有一个个体, 于是

$$p_1(t) = \Pr\{N(t) = 1\} = \Pr\{T > t\} = 1 - F(t),$$

其中

$$F(t) = \Pr\{T \leqslant t\} = \int_0^t f(\tau)\mathrm{d}\tau$$

是 T 的分布函数.

一般地, 令 $G(s,t)$ 是 $N(t)$ 的概率母函数, 即

$$G(s,t) = \sum_{k=0}^{\infty} p_k(t)s^k = \sum_{k=1}^{\infty} p_k(t)s^k.$$

我们将得到关于 $G(s,t)$ 的积分方程. 在时刻 t 恰有 k 个个体的概率 $p_k(t)$ 可用如下方法求得. 第一次分裂以概率 $f(\tau)\mathrm{d}\tau$ 出现在时间区间 $(\tau, \tau + \mathrm{d}\tau)(0 \leqslant \tau \leqslant t)$ 内, 然后两个新的个体分别独立地经历相同的分支过程, 并且在余下时间 $t - \tau$ 内要产生全部 k 个后代. 自然, 第一次分裂时刻 τ 可以取区间 $[0, t]$ 内任何一个值. 于是, 由全概率公式得

$$p_k(t) = \int_0^t \mathrm{d}\tau f(\tau) \sum_{l=1}^{k} p_l(t-\tau)p_{k-l}(t-\tau), \quad k = 2, 3, \cdots,$$

且

$$p_1(t) = 1 - F(t).$$

据 $G(s,t)$ 的定义, 我们有

$$\begin{aligned} G(s,t) &= [1 - F(t)]s + \sum_{k=2}^{\infty} s^k \int_0^t \mathrm{d}\tau f(\tau) \sum_{l=1}^{k} p_l(t-\tau)p_{k-l}(t-\tau) \\ &= [1 - F(t)]s + \sum_{k=0}^{\infty} s^k \int_0^t \mathrm{d}\tau f(\tau) \sum_{l=0}^{k} p_l(t-\tau)p_{k-1}(t-\tau). \end{aligned}$$

既然这里所涉及的量是非负的, 和号与积分号可以互相交换, 从而

432

$$G(s,t)=\int_0^t \mathrm{d}\tau f(\tau)\sum_{k=0}^{\infty}s^k\sum_{l=0}^{k}p_l(t-\tau)p_{k-1}(t-\tau)+[1-F(t)]s.$$

其中和可视为因子 $\sum_{k=0}^{\infty}s^k p_k(t-\tau)=G(s,t-\tau)$ 的卷积. 这样,

$$G(s,t)=\int_0^t [G(s,t-\tau)]^2 f(\tau)\mathrm{d}\tau+[1-F(t)]s. \tag{11.1}$$

不幸的是, 此积分方程一般无法求解. 然而, 当 T 具有密度为

$$f(t)=\lambda\mathrm{e}^{-\lambda t},\quad t\geqslant 0, \tag{11.2}$$

的指数分布时, 我们可以求得其解. 相应于此特殊情况是 Yule 纯生过程, 事实上, 若存在 n 个初始个体, 则至第一次分裂的时间区间长度是随机变量 $Z\min(X_1,X_2,\cdots,X_n)$, 其中 X_i 是独立的且具有分布律 (11.2). Z 服从参数为 $n\lambda$ 的指数分布. 所以, 在以后任一 h 单位时间分裂的机会为 $n\lambda h+o(h)$. 当分裂发生时, 群体所含个体增加至 $n+1$, 并且再下一次分裂的时间区间长度是服从参数为 $(n+1)\lambda$ 的指数分布, 如此等等. 4.1 节已经从纯生过程的观点研究过这个例子. 下面我们用另外的方法来研究.

当 $1-F(t)=\mathrm{e}^{-\lambda t}$ 时, 方程 (11.1) 变为

$$G(s,t)\mathrm{e}^{\lambda t}=\lambda\int_0^t [G(s,t-\tau)]^2\mathrm{e}^{\lambda(t-\tau)}\mathrm{d}\tau+s.$$

记 $u=t-\tau$, 得

$$G(s,t)\mathrm{e}^{\lambda t}=\lambda\int_0^t [G(s,u)]^2\mathrm{e}^{\lambda u}\mathrm{d}u+s$$

关于 t 微分, 导出方程

$$\mathrm{e}^{\lambda t}G'(s,t)+\lambda\mathrm{e}^{\lambda t}G(s,t)=\lambda[G(s,t)]^2\mathrm{e}^{\lambda t},$$

其中

$$G'(s,t)=\frac{\mathrm{d}}{\mathrm{d}t}G(s,t).$$

消去因子 $\mathrm{e}^{\lambda t}$ 之后, 上面方程化为贝努利型微分方程:

$$G'(s,t)=\lambda[G(s,t)]^2-\lambda G(s,t). \tag{11.3}$$

为解此微分方程, 可简单地分离变量:

$$\frac{\mathrm{d}G(s,t)}{G(s,t)[G(s,t)-1]}=\lambda\mathrm{d}t.$$

433

则其解为

$$\frac{G(s,t)-1}{G(s,t)}=C(s)\mathrm{e}^{\lambda t},$$

或简化为

$$G(s,t)=\frac{1}{1-C(s)\mathrm{e}^{\lambda t}},\tag{11.4}$$

其中, $C(s)$ 关于 t 是常数, 但可以是 s 的函数. 为确定 $C(s)$, 在 (11.4) 中令 $t=0$, 由于

$$p_k(0)=\begin{cases}0, & 若\ k\neq 1,\\ 1, & 若\ k=1,\end{cases}\quad s\equiv G(s,0)=\frac{1}{1-C(s)},$$

推得

$$C(s)=\frac{s-1}{s}.$$

故 (11.3) 的解 (也是 (11.1) 在指数分布寿命情况下的解) 是

$$G(s,t)=\frac{s\mathrm{e}^{-\lambda t}}{1-(1-\mathrm{e}^{-\lambda t})s}.\tag{11.5}$$

为求得 $p_k(t)$ 显式表达式, 将 (11.5) 按 s 的幂级数展开, 即

$$G(s,t)=\mathrm{e}^{-\lambda t}s\sum_{k=0}^{\infty}(1-\mathrm{e}^{-\lambda t})^k s^k.$$

显而易见, 我们有

$$p_k(t)=\mathrm{e}^{-\lambda t}(1-\mathrm{e}^{-\lambda t})^{k-1},\quad 对于\ k=1,2,\cdots.$$

虽然在一般情况下无法解积分方程 (11.1), 但我们可由它得到关于均值函数 $m(t)=E[N(t)]$ 的一个等式. 注意

$$\left.\frac{\mathrm{d}G(s,t)}{\mathrm{d}s}\right|_{s=1}=\sum_{k=1}^{\infty}kp_k(t)=m(t).$$

对 s 微分 (11.1) 导致

$$\frac{\mathrm{d}G(s,t)}{\mathrm{d}s}=2\int_0^t G(s,t-\tau)\frac{\mathrm{d}G(s,t-\tau)}{ds}f(\tau)\mathrm{d}\tau+1-F(t).$$

两边令 $s=1$, 由于

434

$$G(1,t-\tau)=\sum_{k=1}^{\infty}p_k(t-\tau)=1.$$

则

$$m(t)=2\int_0^t m(t-\tau)f(\tau)\mathrm{d}\tau+1-F(t).$$

此积分方程可作为更新方程的一个例子. 它的特点是未知函数以卷积形式出现于积分号内. 有许多经典理论涉及到更新方程. 它描述了解 $m(t)$ 的渐近增长性质.

上述模型的一个自然推广是允许每个个体在其生命结束时恰好分裂成 r 个新的同类型个体, 其中 r 是比 2 大的固定整数. 此时易见, 积分方程 (11.1) 变为

$$G(s,t)=\int_0^t [G(s,t-\tau)]^r f(\tau)\mathrm{d}\tau+[1-F(t)]s.$$

这个模型的进一步推广是每个个体在其生命结束时分裂成随机数目的同类型新个体. 例如, 我们可以假设每个个体死亡时以概率 $q_l(l=0,1,2,\cdots,)$ 产生 l 个新的同类型个体. 令

$$h(s)=\sum_{l=0}^{\infty}q_l s^l$$

是相应的母函数. 我们可用下述方法导出积分方程. 设第一次分裂出现在时刻 $\tau(0\leqslant\tau\leqslant t)$ 且有 l 个新的个体产生. 这个事件的概率为 $f(\tau)\mathrm{d}\tau q_1$ 在余下 $t-\tau$ 个单位时间内, l 个个体的每一个都能产生后代使得在时刻 t 群体含量达到 k. 依全概率公式, 我们有

$$p_k(t)=\int_0^t \mathrm{d}\tau f(\tau)\sum_{l=0}^{\infty}q_l\sum_{k_1+k_2+\cdots+k_l=k}p_{k_1}(t-\tau)p_{k_2}(t-\tau)\cdot\cdots\cdot p_{k_l}(t-\tau)$$

$$\text{对于}k=2,3,\cdots\quad \text{和}p_1(t)=1-F(t).$$

母函数为

$$\begin{aligned}G(s,t)&=[1-F(t)]s+\sum_{k=2}^{\infty}s^k\int_0^t \mathrm{d}\tau f(\tau)\sum_{l=0}^{\infty}q_l\sum_{k_1+\cdots+k_l=k}p_{k_1}(t-\tau)\cdots p_{k_l}(t-\tau)\\&=[1-F(t)]s+\int_0^t \mathrm{d}\tau f(\tau)\sum_{l=0}^{\infty}q_l\sum_{k=1}^{\infty}s^k\sum_{k_1+\cdots+k_l=k}p_{k_1}(t-\tau)\cdots p_{k_l}(t-\tau).\end{aligned}$$

435

但

$$\sum_{k=0}^{\infty}s^k\sum_{k_1+\cdots+k_l=k}p_{k_1}(t-\tau)\cdots p_{k_l}(t-\tau)$$

的内和是 l 重卷积. (学生应自行证明.) 其中每因子对应下式

$$\sum_{k=0}^{\infty} s^k p_k(t-\tau) = G(s, t-\tau),$$

总和为 $[G(s,t-\tau)]^l$. 故有

$$G(s,t) = [1-F(t)]s + \int_0^t \mathrm{d}\tau f(\tau) \sum_{l=0}^{\infty} q_l [G(s,t-\tau)]^l.$$

等式右端积分中的和式给出了在 $x = G(s,t-\tau)$ 处的母函数 $h(s)$. 在此情况下, 最终积分等式的形式如下:

$$G(s,t) = \int_0^t h(G(s,t-\tau)) f(\tau) \mathrm{d}\tau + [1-F(t)]s.$$

初等问题

1. 设 X_n 是一分支过程且 $X_0 \equiv 1$. 对任给固定正整数 k, 定义序列

$$Y_r = X_{rk}, r = 0, 1, 2, \cdots$$

试证 $\{Y_r, r = 0, 1, 2, \cdots\}$ 构成一分支过程, 若 $\varphi(s)$ 是过程 X_n 中单一个体直接后代随机变量的母函数, $\varphi_n(s)$ 是其第 n 次迭代式, 则 $\varphi_k(s)$ 是过程 Y_r 中单一个体直接后代随机变量的母函数.

2. 设 $f(s) = 1 - p(1-s)^\beta$, 其中 p 和 β 是常数且 $0 < p < 1$, $0 < \beta < 1$. 试证明 $f(s)$ 是概率母函数且它的迭代式为

$$f_n(s) = 1 - p^{1+\beta+\cdots+\beta^{n-1}} (1-s)^{\beta^n}, \quad n = 1, 2, \cdots.$$

3. 设 $f(s)$ 是概率母函数, $h(s)$ 是满足

$$g(s) = h^{-1}[f(h(s))]$$

的一个概率母函数, 试验证

$$g_n(s) = h^{-1}[f_n(h(s))]$$

是一个概率母函数, 其中 f_n 和 g_n 分别表示函数 f 和 g 的第 n 次迭代函数式. [436]

4. 作为初等问题 3 的一个例子, 取

$$f(s) = \frac{s}{m-(m-1)s} \quad (m > 1)$$

和

$$h(s) = s^k \quad (k \text{ 为正整数}).$$

试证 $g(s)=h^{-1}f(h(s))$ 是母函数, 且 g 的第 n 次迭代式为

$$g_n(s)=\frac{s}{(m^n-(m^n-1)s^k)^{1/k}}.$$

5. 证明在分支过程 X_n 中若 $m=E(X_1)<1$ 则

$$E\left(\sum_{n=1}^{\infty}X_n\right)=m/(1-m).$$

6. 设从一红细胞开始进行血培养, 一分钟之后红细胞死亡并按下面所指定的概率产生对应的一组细胞.

2 红细胞	$\frac{1}{4}$,
1 红细胞, 1 白细胞	$\frac{2}{3}$,
2 白细胞	$\frac{1}{12}$.

每个红细胞生活一分钟之后又以同样方式产生后代细胞, 而每个白细胞生活一分钟死亡之后不产生后代. 假定各个体细胞之间是相互独立的. 试求

(a) 血培养开始之后 $n+\frac{1}{2}$ 分钟不出现白细胞的概率.

(b) 血培养过程最终完全死亡的概率.

答案: (a) $\left(\frac{1}{4}\right)^{2^n-1}$; (b) $\frac{1}{3}$.

7. 设 $f(s)=as^2+bs+c$, 其中 a,b,c 是正的且 $f(1)=1$. 假设消失概率为 $d(0<d<1)$. 试证明

$$d=\frac{c}{a}.$$

8. 设某分支过程一个初始粒子后代个数的分布具有母函数 $f(s)$. 第一代每个成员后代个数的分布具有母函数 $g(s)$. 第二代每个成员后代个数的分布具有母函数 $f(s)$. 再下一代母函数又是 $g(s)$. 母函数如此一代代更替下去.

试由基本原理 (即不利用 8.5 节和 8.6 节多维理论一般结果) 确定该过程的消失概率和第 n 代粒子的平均数. 若过程初始母函数为 $g(s)$. 然后以同样方式更替, 所求这些量是否改变?

9. 对于任一离散时间分支过程 $X_n, X_0=1$, 证明不等式

437

$$\Pr\{X_n>L \text{ 对于某些 } 0\leqslant n\leqslant m|X_m=0\}\leqslant[\Pr\{X_m=0\}]^L.$$

10. 考虑一分支过程, 其初始时刻含有 N 个个体, 概率母函数为

$$\varphi(s)=q+ps,\quad q,p>0,q+p=1.$$

试确定群体消失的首达时间 T 的概率分布.

答案: $\Pr\{T=n\}=(1-p^n)^N-(1-p^{n-1})^N$.

11. 计算 $\mathrm{Var}[X(t)]$, 其中 $X(t)$ 是连续时间分支过程且 $X(0)=1$.

答案:

$$\operatorname{Var}[X(t)]=\begin{cases}\left[\dfrac{u''(t)-u'(1)}{u'(1)}\right]\mathrm{e}^{u'(1)t}(\mathrm{e}^{u'(1)t}-1), & 若\ u'(1)\neq 0\\ u''(1)t & 若\ u'(1)=0\end{cases}$$

12. 设一群体是由雌的和雄的两种类型个体组成. 如果群体中至少有一个雄性, 则所有雌性能够产生后代, 其后代数目的母函数为 $f(x)$. 若一后代是雌的概率为 α, 并已知至少也有一雄性产生, 试求下一代雌性数目的概率母函数.

答案: $\dfrac{f(\alpha s+(1-\alpha))-f(\alpha s)}{1-f(\alpha)}$.

问 题

1. 为研究泌尿过程引入下面模型. 假设细菌增长依照参数 λ 的 Yule 过程 (见 4.1 节). 在每单位时间, 现存的每一个细菌以概率 p 消灭. 试求在时刻 n 时细菌存留数的概率母函数.

提示: 这是分支过程第 n 代的母函数.

答案: $f_n(s)$ 是下面函数的第 n 次迭代

$$f(s)=\frac{\mathrm{e}^{-\lambda}(p+qs)}{1-(1-\mathrm{e}^{-\lambda})(p+qs)}.$$

2. (a) 一成熟个体依概率母函数 $f(s)$ 产生后代. 假设一群体有 k 个不成熟个体, 每个以概率 p 成熟, 然后独立地繁殖其他个体. 试求下一代 (不成熟) 个体数的概率母函数.

(b) 若已知现在这一代有 k 个成熟个体, 试求下一代成熟个体数的概率母函数. 438

答案: (a) $(1-p+pf(s))^k$; (b) $(f(1-p+ps))^k$.

3. 证明在问题 2 中, (a) 和 (b) 的分布有相同均值但不一定有相同方差.

4. 考虑一离散时间分支过程 $\{X_n\}$, 其概率母函数为

$$\varphi(s)=\frac{1-(b+c)}{1-c}+\frac{bs}{1-cs},\quad 0<c<b+c<1,$$

其中 $(1-b-c)/c(1-c)>1$. 设 $X_0=1$. 试确定条件极限分布:

$$\lim_{n\to\infty}\Pr\{X_n=k|X_n>0\}.$$

答案: $\left(1-\dfrac{1}{s_0}\right)\left(\dfrac{1}{s_0}\right)^{k-1},\quad s_0=\dfrac{1-b-c}{c(1-c)}$.

5. 在上一问题中假设 $1-b-c=c(1-c)$. 试确定 $\Pr\{X_n>0\}$.

答案: $(1-c)/[1+(n-1)c]$.

6. 在问题 5 相同条件下, 证明 $\Pr\{X_n\leqslant nx|X_n>0\}$ 收敛于一指数分布.

提示: 在条件 $x_n>0$ 之下计算 X_n/n 的拉普拉斯变换, 并求其当 $n\to\infty$ 时的极限.

答案: 参数为 $(1-c)/c$ 的指数分布.

7. 设一连续时间分支过程具有无限小母函数

$$u(s) = s^k - s \quad (k \text{ 为不小于 2 的整数}).$$

试求其母函数 $\varphi(t;s)$.

提示：解

$$\frac{\partial\varphi(t;s)}{\partial t} = u(\varphi(t;s)), \varphi(0;s) = s$$

答案：$\varphi(t,s) = s[\mathrm{e}^{(k-1)t} - (\mathrm{e}^{(k-1)t} - 1)s^{k-1}]^{-1/(k-1)}$.

8. 若连续时间分支过程的无限小母函数是

$$u(s) = 1 - s - \sqrt{1-s},$$

试求其母函数 $\varphi(t;s)$.

答案：$\varphi(t;s) = 1 - [1 - \mathrm{e}^{-t/2} + \mathrm{e}^{-t/2}\sqrt{1-s}]^2$.

9. 考虑多生 Yule 过程, 群体的每个成员在长度为 h 的时间区间内以概率 $\beta h + o(h)$ 产

439

生 k 个新的成员, 并以概率 $(1 - \beta h + o(h))$ 不产生新的成员 $(\beta > 0, k$ 为正整数). 设在时刻 0 有 N 个成员.

(a) 设 $X(t)$ 是至时刻 t 分裂的数目, 试确定 $E(X(t))$ 增长的状态.

(b) 设 τ_n 是第 n 次分裂时间, 求 τ_n 的密度函数.

提示：(a) 注意

$$\Pr\{\tau_n \leqslant t | \tau_{n-1} = \xi\} = \begin{cases} 1 - \exp\{-[(n-1)k + N]\beta(t-\xi)\}, & \xi \leqslant t \\ 0, & \xi > t \end{cases}$$

并得到用 τ_{n-1} 密度函数表示的 τ_n 密度函数的递归公式.

答案：

(a) $\dfrac{E[X(t)]}{\mathrm{e}^{k\beta t}} \to \dfrac{N}{k}, \quad t \to \infty.$

(b) 记 τ_n 的概率密度函数为 $f_n(t)$, 则

$$f_n(t) = \frac{N(N+k)\cdots[N+(n-1)k]}{(n-1)!k^{n-1}}\beta\mathrm{e}^{-N\beta t}(1 - \mathrm{e}^{-k\beta t})^{n-1}.$$

10. 设 $X_n, n \geqslant 0$, 是概率母函数为 $\varphi(s)$ 的分支过程, 其前 n 代个体总数. 定义为 Y_n, 即

$$Y_n = X_0 + X_1 + \cdots + X_n, \quad n = 0, 1, 2, \cdots, X_0 = 1,$$

设 $F_n(s)$ 是 Y_n 的概率母函数, 试证明

$$F_{n+1}(s) = s\varphi(F_n(s)), \quad n = 0, 1, 2, \cdots.$$

11. 设一分支过程在初始时刻仅有一个个体, 其单个个体后代数目的母函数为 $\varphi(s)$, $\varphi(s)$ 的第 n 次迭代记为 $\varphi_n(s)$. 此外, 每代有个体迁入, 其概率母函数为 $h(s)$. 考虑有迁入的分支过程, 其转移概率矩阵被下式所确定：

$$\sum_{j=0}^{\infty} P_{ij}s^j = [\varphi(s)]^i \cdot h(s).$$

试证明其第 n 步转移概率矩阵被下式所确定:

$$\sum_{j=0}^{\infty} P_{ij}^{n} s^{j} = [\varphi_n(s)]^{i} h(\varphi_{n-1}(s)) h(\varphi_{n-2}(s)) \cdot \cdots \cdot h(\varphi(s)) h(s).$$

12. 在有迁入的分支过程中 (问题 11), 假定 $\varphi'(1) = m < 1$. 试证明其对应的马尔可夫链具有母函数为 $\pi(s) = \sum_{r=0}^{\infty} \pi_r s^r$ 的平稳概率分布, 并且 $\pi(s)$ 满足函数方程

440

$$\pi(\varphi(s)) h(s) = \pi(s).$$

13. 在问题 12 中, 若指定 $\varphi(s) = q + ps (0 < p < 1, q + p = 1)$ 和 $h(s) = \mathrm{e}^{s-1}$, 试确定其平稳概率分布.

14. 考虑简单生灭过程 (无迁入线性增长过程), 即 $\lambda_n = \lambda n$ 和 $\mu_n = \mu n$, 其中 $\lambda > 0, \mu > 0$ 且 $\mu > \lambda$. 设 $Z(t)$ 是时刻 t 群体的含量. 此外, 有一无限服务的排队过程, 顾客到达间隔时间分布为 $1 - \mathrm{e}^{-\lambda t}$ 而服务时间分布为 $1 - \mathrm{e}^{-\mu t}$. 试证明该排队过程的忙碌期和在初始条件 $Z(0) = 1$ 下的 $\int_0^{\infty} Z(t) \mathrm{d}t$ 具有相同的分布.

15. 设 X_n 为离散分支过程, 其对应的概率母函数为 $\varphi(s)$, 令 $\varphi_n(s) = \sum_{k=0}^{\infty} \Pr\{X_n = k\} s^k$, 并设 $\varphi'(1) > 1$. 现令 $\widetilde{X}_n$ 表示具有无限世代的第 n 代粒子总数. 试证 $\widetilde{X}_n$ 的概率母函数是

$$\sum_{k=0}^{\infty} \Pr\{\widetilde{X}_n = k | \widetilde{X}_0 = X_0 = 1\} s^k = \frac{\varphi_n(s(1-q)+q) - q}{1-q}$$

其中 q 是消失概率.

提示: 注意, 对于 $k \geqslant 1$ 有

$$\Pr\{\widetilde{X}_n = k | \widetilde{X}_0 = 1, X_0 = 1\} = \frac{\sum_{l=k}^{\infty} \Pr\{\widetilde{X}_n = k, X_n = l | X_0 = 1\}}{\Pr\{\widetilde{X}_0 = 1 | X_0 = 1\}}.$$

16. 下面问题目的在于确定不同形式的死亡对群体稳定性的影响. 所谓稳定性定义为无限残存的概率 =1− 最终消失的概率.

我们暂时考虑无另外死亡情况. 单独个体后代数 X 具有概率分布

$$\Pr\{X = k\} = p_k, \quad k = 0, 1, \cdots.$$

设分布均值为 m, 群体中所有后代相互独立且同分布.

下面考虑三种类型死亡方式. 在每种情况下, 一个体幸存的概率均为 p, 但各取不同形式. 设 $mp > 1$.

(a) 个别死亡: 其发生独立于其他个体, 每个个体幸存的概率为 p. 换句话说, 若给定一个体的后代数 X(例如一胎生下的动物数), 其实际存活后代个体数服从参数为 (X, p) 的二项分布. 这种形式死亡可能出现于成年捕食性动物.

441

(b) 胎死, 其发生独立于别的胎, 每胎以概率 p 存留或以概率 $q = 1 - p$ 死亡. 即, 若给定个体数为 X 的一胎, 则这一胎实际数以概率 p 为 X, 以概率 q 为 0. 这种形式死亡可能出现于幼年捕食性动物或鸟类的雏和蛋.

(c) 整代死亡: 整代以概率 p 幸存或以概率 q 全部死亡. 这种类型的死亡可能出现于诸如森林火灾, 洪水等自然灾害的影响.

每种情况下试确定 1− 稳定性 =Pr{最终消失}的方程式, 由此回答哪一类群体是最稳定的? 哪一类是最不稳定的? 您能否证之?

附 记

本章写作的灵感源于 T. Harris 关于分支过程的专题论文 [1]. [1] 还包含了有关分支过程及其应用的丰富文献目录.

参考书目

442 [1] T. Harris, *The Theory of Branching Processes.* Springer-Verlag, Berlin, 1963.

第9章 平稳过程

平稳过程是概率规律不随时间 (有时是空间) 改变而变化的随机过程. 这个概念描述了无固有时间起点 (空间原点) 的物理系统的自然特性. 对于诸如通讯理论、天文学、生物学、生态学以及经济学中出现的大量过程, 这种假设是适当的.

平稳性在理论上导致了许多重要结论. 本章我们致力于预测问题、遍历性、平稳过程的谱表示、平稳点过程以及水平交叉问题. 9.1 节和 9.2 节是后面各节的预备知识. 9.3 节和 9.4 节是关于预测理论, 9.5 节和 9.6 节是关于遍历理论, 9.7 节和 9.8 节关于谱分析, 9.9 节和 9.10 节关于点过程和水平交叉问题, 这些内容读者可根据需要按任意顺序阅读.

9.1 定义和例子

设 T 是抽象指标集, 且 T 中任意两点之和仍然属于 T. 通常 T 是非负整数集合 $\{0,1,\cdots\}$, 但它也可以是整条直线、正半轴、平面、有限维空间、球面, 或者甚至是无限维空间.

定义 1.1 平稳过程是这样一个随机过程 $\{X(t), t\in T\}$, 对于任何正整数 k 和 T 中任意点 $t_1,\cdots,t_k$ 及 h,

$$\{X(t_1),\cdots,X(t_k)\}$$

的联合分布与

$$\{X(t_1+h),\cdots,X(t_k+h)\}$$

的联合分布相同.

下面是一些简单例子:

(a) 在通信理论中电子脉冲通常假设是一个平稳过程. 当然, 在任何物理系统中, 在信号的开始都有一个过渡时间. 由于这个过渡时间比之信号长度是很短的时间, 因而平稳模型是适用的. 在电子通讯理论中, 电位和电流通常都用复变量来表示. 这时, 我们会遇到复值平稳过程.

(b) 星球和星系的空间分布, 植物和动物的平面分布, 常常是平稳的. 这里 T 可以是欧几里得空间. 球面或者平面.

对于海浪的高度也可以假设是平稳分布, T 取作经纬度, 仍然是二维的.

(c) 经济时间序列, 例如失业人数, 国民生产总值, 国民收入等等, 通常假设对应于一个平稳过程, 至少对已经实现的长期增长作某种修正之后是这样一个过程.

如上述例子所说明的, 平稳过程以各种各样形式大量出现. 处理最一般的情况与本书仅提供一个导引的初衷相违背. 已经提醒读者注意广泛的范围, 下面我们主要集中在实值过程并且 $T=\{0,1,2,\cdots\}$ 的简单情况.

设 $\{X(t),t\in T\}$ 是平稳过程. 如果均值 $m(t)=E[X(t)]$ 存在, 则这个量必须是常数, 即对所有 $t, m(t)=m$. 类似地, 如果二阶矩 $E[X(t)^2]$ 是有限的, 则方差 $\sigma^2=E[(X(t)-m)^2]$ 是与时间无关的常数. 设 t 和 s 是两个时刻, 不失一般性, 假设 $t>s$. 利用平稳性, 我们计算协方差

$$E[(X(t)-m)(X(s)-m)]=E[(X(t-s)-m)(X(0)-m)],$$

其右端仅仅依赖于差 $t-s$. 如果我们定义协方差函数

$$R(h)=E[(X(h)-m)(X(0)-m)],$$

则

$$E[(X(t)-m)(X(s)-m)]=R(|t-s|).$$

自然, $\sigma^2=R(0)$. 把协方差标准化便得到所谓的**相关函数**或**自相关函数**, 其定义如下

444

$$\rho(u)=\frac{1}{\sigma^2}R(u)=R(u)/R(0).$$

显然 $\rho(0)=1$. 可以证明 (利用施瓦兹不等式) 对于所有的 u 有 $-1\leqslant\rho(u)\leqslant 1$.

定义 1.1 给出的平稳概念包含了过程的所有有限维的分布. 出于种种目的, 我们希望有一个实用的比较弱的概念, 它仅仅包含前两阶矩.

定义 1.2 一个随机过程 $\{X(t),t\in T\}$ 称为**协方差平稳过程**, 如果它的二阶矩有限, $E[X(t)^2]<\infty$, 均值为常数 $m=E[X(t)]$, 并且协方差 $E[(X(t)-m)(X(s)-m)]$ 仅依赖于时间差 $|t-s|$.

在文献中还使用与协方差平稳同义的术语是**弱平稳**或**宽平稳**. 为强调区别, 定义 1.1 引入的平稳过程常常称为**严格平稳的**.

具有有限二阶矩的平稳过程是协方差平稳的 (当然, 平稳过程可能没有有限矩). 一个协方差平稳过程很可能不是平稳的, 但有一个重要例外. 一个随机过程 $\{X(t),t\in T\}$, 对于任意 k 和每个有限时间集合 $\{t_1,t_2,\cdots,t_k\}$, 随机向量

$$(X(t_1),\cdots,X(t_k))$$

具有多元正态分布 (第 1 章), 则称为**高斯过程**. 由于多元正态分布是由它的前两阶矩所决定, 即由均值向量和协方差矩阵所决定, 因此协方差平稳的高斯过程必定是严格平稳过程.

例子

下面几个例子将在以后进一步加以讨论.

A. 两个极端的平稳过程

(i) 独立同分布随机变量序列 $Y_0, Y_1, \cdots$ 是平稳过程. 如果 $Y_0, Y_1, \cdots$ 的共同分布具有有限方差 σ^2, 则过程是协方差平稳的, 协方差函数是

$$R(u) = \begin{cases} \sigma^2, & \text{对于} u = 0, \\ 0, & \text{对于} u \neq 0. \end{cases}$$

445

(ii) 考虑另一个完全不同的平稳过程, 令 Z 是具有已知分布的单一随机变量, 置 $Z_0 = Z_1 = Z_2 = \cdots = Z$. 显而易见, 过程 $\{Z_n\}$ 是平稳的. 如果随机变量 Z 具有有限方差, 则过程是协方差平稳的, 并且协方差函数

$$R(u) = \sigma^2, \quad \text{对于所有的} u.$$

从许多方面来说, $\{Y_n\}$ 和 $\{Z_n\}$ 是两个极端的例子, 它们可以用来说明平稳过程性质的极大差异. 例如, 已知共同分布, 观测 $Y_0, Y_1, \cdots, Y_n$ 对预报 Y_{n+1} 不会提供任何信息, 而只要观测 Z_0 就能准确预报 $Z_1, Z_2, \cdots$. 另外, 我们还可以从另一方面说明上述两个过程是相反的. 假设 Y_n 具有有限均值, 利用大数定理, 样本平均值

$$\frac{1}{n}(Y_0 + \cdots + Y_{n-1})$$

收敛于常数 $m = E[Y_0]$. 但对于过程 $\{Z_n\}$, 则没有这样的收敛性. 事实上

$$\frac{1}{n}\{Z_0 + \cdots + Z_{n-1}\} = Z_0 = Z,$$

即 n 个样本平均值的"随机性"与第一个观测时是一样的.

由过程构成的样本平均值收敛于过程的某个基本参数, 这样的性质称为是**遍历的**. 对遍历过程的基本参数作出推断, 并不需要对整个过程或样本轨道进行反复多次的独立观察. 相反, 我们只需要观测过程的一个单一实现, 不过要有足够长的时间. 因此在实际中一个重要课题是确定在什么条件下平稳过程是遍历的.

令人意外的是, 这两个性质相反的例子却具有相关联的起因. 对于协方差平稳过程, 这两种情况问题症结都在于当时间差 $|t-s|$ 变大时协方差函数 $R(|t-s|)$ 是否收敛于 0, 如果它收敛于 0, 其收敛速度为何. 对于过程 $\{Y_n\}$, 收敛速度是非常快的, 因为当时间差不小于 1 时协方差恰为 0, 而对于过程 $\{Z_n\}$ 不论时间差取什么值相关函数都是 1.

446

平稳过程理论的基本目的是阐述介于上述两个极端例子之间的大量平稳过程的遍历性质和预测问题. 平稳过程的一般预测问题在本章的 9.3 节和 9.4 节中研究, 样本平均值的收敛性在 9.5 节和 9.6 节中详细叙述.

B. 三角多项式

平稳过程的某些有趣例子可以由具有随机振幅三角表达式得到. 设 A 和 B 是均值为 0、方差为 σ^2 的同分布随机变量, 我们假设 A 和 B 是不相关的, 即 $E[AB]=0$. 固定一个频率 $\omega \in [0,\pi]$ 并对 $n=0,\pm 1,\pm 2,\cdots$ 定义

$$X_n = A\cos(\omega n) + B\sin(\omega n).$$

则对于所有的 $n, E[X_n]=0$, 利用三角恒等式

$$\cos(\alpha-\beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta,$$

以及 $E[AB]=0$, 我们求得协方差为

$$\begin{aligned} &E[X_nX_{n+u}] \\ &= E\left[\{A\cos\omega n + B\sin\omega n\}\{A\cos\omega(n+u) + B\sin\omega(n+u)\}\right] \\ &= E\left[A^2\cos\omega n\cos\omega(n+u) + B^2\sin\omega n\sin\omega(n+u)\right] \\ &= \sigma^2\cos\omega u. \end{aligned}$$

由于 X_n 和 X_{n+u} 的协方差显然仅依赖于时间差 u, 我们断定过程是协方差平稳的. 如果 A 和 B 具有均值为 0 和方差为 σ^2 的正态分布, 则过程是高斯的, 因而是严格平稳的.

特别地, 对于频率 $\omega=0$, 我们有 $\cos\omega n = 1$ 和 $\sin\omega n = 0$, 所以对于所有的 $n, X_n = A$. 这样, 前面例子中的过程 $\{Z_n\}$ 在这里就出现了.

更加一般地, 令 $A_0, A_1, \cdots, A_m$ 和 $B_0, B_1, \cdots, B_m$ 是具有零均值的, 且不相关的随机变量. 假设 A_i 和 B_i 具有共同的方差 σ_i^2, 并令 $\sigma^2 = \sigma_0^2 + \cdots + \sigma_m^2$. 取 ω_0, $\omega_1, \cdots, \omega_m$ 为 $[0,\pi]$ 中不同的频率, 并对于 $n=0,\pm 1,\pm 2,\cdots$ 置

447

$$X_n = \sum_{k=0}^{m}\{A_k\cos n\omega_k + B_k\sin n\omega_k\}.$$

由于系数 $\{A_k\}$ 和 $\{B_k\}$ 不相关且均值为 0, 我们有 $E[A_iB_j]=0$ 和 $E[A_iA_k] = E[B_iB_k] = 0, k \neq i$. 我们计算协方差

$$\begin{aligned} E[X_nX_{n+u}] &= E\left[\left(\sum_{k=0}^{m}\{A_k\cos n\omega_k + B_k\sin n\omega_k\}\right)\left(\sum_{j=0}^{m}\{A_j\cos(n+u)\omega_j\right.\right. \\ &\qquad \left.\left. + B_j\sin(n+u)\omega_j\}\right)\right] \\ &= \sum_{k=0}^{m} E[A_k^2\cos n\omega_k\cos(n+u)\omega_k + B_k^2\sin n\omega_k\sin(n+u)\omega_k] \end{aligned}$$

$$= \sum_{k=0}^{m} \sigma_k^2 \cos u\omega_k.$$

因此, 此过程也是协方差平稳的.

继而令 $p_k = \sigma_k^2/\sigma^2$, 记协方差函数为

$$R(u) = \sigma^2 \sum_{k=0}^{m} p_k \cos u\omega_k. \tag{1.1}$$

则 p_k 表示频率 ω_k 在协方差中的贡献. 注意到 $\{p_k\}$ 是离散概率分布, 即 $p_k \geqslant 0$ 和 $\sum p_k = 1$, 由此提示我们把 (1.1) 推广到频率连续的情形, 其形式为

$$R(u) = \sigma^2 \int_0^{\pi} \cos(u\omega)\mathrm{d}F(\omega), \tag{1.2}$$

其中 $F(\omega)$ 是取值于 $[0,\pi]$ 的随机变量的分布函数. 在 9.7 节中, 我们将看到这样的推广确实是可能的, 并且最一般的协方差平稳过程都具有这种表达式. 在 F 对应于 $[0,\pi]$ 上均匀分布的特殊情况下, 意味着所有频率出现的可能性是相同的. 我们计算

$$R(u) = \sigma^2 \frac{1}{\pi} \int_0^{\pi} \cos(u\omega)\mathrm{d}\omega$$

$$= \begin{cases} \sigma^2, & 如果 \quad u = 0, \\ 0, & 如果 \quad u \neq 0. \end{cases}$$

这是前述例子中独立同分布序列 $\{Y_n\}$ 的协方差函数. 这里再一次体现出 $\{Y_n\}$ 和 $\{Z_n\}$ 具有某种意义的相对性, $\{Z_n\}$ 对应于单一频率 $\omega = 0$, 而 $\{Y_n\}$ 对应于在$[0,\pi]$ 均匀分布的频率. 448

C. 滑动平均过程

令 $\{\xi_n : n = 0, \pm 1, \pm 2, \cdots\}$ 是具有共同均值 μ 和方差 σ^2 的不相关随机变量序列. $a_1, a_2 \cdots, a_m$ 是任意实数, 考虑由下式定义的过程

$$X_n = a_1\xi_n + a_2\xi_{n-1} + \cdots + a_m\xi_{n-m+1}.$$

我们有

$$E[X_n] = \mu(a_1 + \cdots + a_m)$$

和

$$\mathrm{Var}[X_n] = \sigma^2(a_1^2 + \cdots + a_m^2).$$

记 $\widehat{\xi}_k = \xi_k - \mu$. 计算其协方差

$$E\left[\left(X_n - \mu\sum_{i=1}^{m} a_i\right)\left(X_{n+u} - \mu\sum_{i=1}^{m} a_i\right)\right]$$

$$
\begin{aligned}
&= E\left[\left(\sum_{i=1}^{m} a_i\widehat{\xi}_{n-i+1}\right)\left(\sum_{j=1}^{m} a_j\widehat{\xi}_{n+u-j+1}\right)\right] \\
&= \begin{cases} E[a_m a_{m-v}\widehat{\xi}^2_{n+u-m+1} + a_{m-1}a_{m-u-1}\widehat{\xi}^2_{n+u-m+2} \\ \qquad\qquad + \cdots + a_{v+1}a_1\widehat{\xi}^2_n], & 若 \quad u \leqslant m-1, \\ 0, & 若 \quad u \geqslant m, \end{cases} \\
&= \begin{cases} \sigma^2(a_m a_{m-u} + \cdots + a_{u+1}a_1), & 若 \quad u \leqslant m-1, \\ 0, & 若 \quad u \geqslant m. \end{cases}
\end{aligned}
$$

由于 X_n 和 X_{n+u} 的协方差仅依赖于延迟时间 u, 而并不依赖于 n, 所以过程是协方差平稳的.

通常的情况是具有标准方差和 $a_k = 1/\sqrt{m}, (k = 1, \cdots, m)$ 的 "滑动平均". 其协方差函数变为

$$
R(u) = \begin{cases} \sigma^2\left(1 - \dfrac{u}{m}\right), & u \leqslant m-1, \\ 0, & u \geqslant m. \end{cases}
$$

当 $m = 1$ 时对应于例 A 中的不相关随机变量序列 $\{Y_n\}$. 粗略地说, 另一个极端, $m = \infty$ 对应于过程 $\{Z_n\}$.

449

D. 圆周上的平稳过程

设 U, V 是具有零均值和单位方差的独立正态分布随机变量. 令 $T = [0, 2\pi]$, 对于 $t \in T$ 定义一个双变量过程 $X(t) = (Y(t), Z(t))$, 其中

$$
\begin{aligned}
Y(t) &= U\sin t + V\cos t, \\
Z(t) &= -U\cos t + V\sin t,
\end{aligned}
$$

则 $(X(t), t \in T)$ 是平稳过程. 显然, $E[Y(t)] = E[Z(t)] = 0$, 并且对于所有的 t 有 $\sin^2 t + \cos^2 t = 1$, 故 $E[Y(t)^2] = E[Z(t)^2] = 1$. 由于 $Y(t)$ 和 $Z(t)$ 具有联合正态分布, 为了完全确定它们的分布, 我们只要计算它们的协方差. 但是, 显然

$$
E[Y(t)Z(t)] = 0.
$$

这样, $X(t)$ 的分布与 $X(t+\theta)$ 的分布相同, 其中 θ 是任意的. 对于任意向量 $(X(t_1), X(t_2), \cdots, X(t_k))$ 我们可以验证相同的性质, 所以过程是平稳的.

E. 平稳马尔可夫链

在第 3 章我们曾经证明在很一般的条件下, 一个马尔可夫链 $\{X_n\}$ 渐趋于一个平衡的统计波动系统, 而时间 n 和初始状态 X_0 最后将不起什么作用. 更加精确地

说, 在我们给出的条件下, 当 n 变大时, 分布 $\Pr\{X_n = j|X_0 = i\}$ 将逼近于一个不依赖于 i 的常数 π_j. 现在假设我们正在观察这样一个系统, 它从无穷远的过去开始发展且现在正处于平衡状态. 这样对任何 $n = 0, \pm1, \pm2, \cdots, X_n$ 的概率分布不依赖于 n. 确实, $\Pr\{X_n = j\} = \pi_j$, 所以这个概率, 或 X_n 的边缘分布在不依赖于 n 的意义下是平稳的. 类似地, (X_n, X_{n+1}) 的联合分布也不依赖于 n, 它可以由下式给出

$$\Pr\{X_n = i, X_{n+1} = j\} = \Pr\{X_n = i\}\Pr\{X_{n+1} = j \Big| X_n = i\} = \pi_i P_{ij},$$

其中 $\boldsymbol{P} = \|P_{ij}\|$ 是马尔可夫链的转移概率矩阵. 显然, 我们可以继续说明对任意固定的 $k = 1, 2, \cdots$, 联合分布 $(X_n, X_{n+1}, \cdots, X_{n+k})$ 不依赖于 n.

类似地, 令 $\{X_n, n = 0, 1, \cdots\}$ 是马尔可夫链, 现按照平稳分布 π_j 选取其初始状态 X_0, 同理可证 $\{X_n\}$ 是一个平稳过程.

450

9.2 平均平方距离

由于 "协方差平稳" 是一个仅利用随机过程前两阶矩定义的性质, 因此我们也希望仅利用前两阶矩来度量任意两个随机变量 Y, Z 的差异或 "距离". 对这样一个度量的自然选择就是平均平方差 $E[(Y-Z)^2]$, 或者平均平方差的平方根, 后者的好处就是与原来的度量单位保持一致. 为了更好地表达这个距离, 引进记号

$$\|Z\| = \sqrt{E[Z^2]} \quad 和 \quad \|Y - Z\| = \sqrt{E[(Y-Z)^2]}.$$

这样, 我们就可以用平均平方差的平方根 $\|\widehat{X}_{t+k} - X_{t+k}\|$ 来度量 $\widehat{X}_{t+k}$ 作为随机变量 X_{t+k} 的预测的能力. 另外, 利用这个距离度量我们可导出随机变量序列相应的收敛性概念. 一个随机变量序列 $\{X_n\}$ 称为收敛于随机变量 X, 如果平均平方差 $E[(X_n - X)^2]$ 当 n 无限增大时趋于 0, 其正式定义如下:

定义 2.1 设 $X, X_1, X_2, \cdots$ 是一列随机变量. 我们说 $\{X_n\}$ 平均平方收敛于 X, 记为 $X_n \to X(\text{m.s.})$ 或 $\lim\limits_{n\to\infty} X_n = X(\text{m.s.})$, 如果

(i) 对于所有的 $n, E[X_n^2] < \infty$,

(ii) $\lim\limits_{n\to\infty} E[(X_n - X)^2] = 0$(或等价地, $\lim\limits_{n\to\infty} \|X_n - X\| = 0$).

下面是平均平方距离 (以下简称均方距离) 和平均平方收敛 (以下简称均方收敛) 的一些初等性质. 其中, Y, Z 是任意两个具有有限二阶矩的随机变量, 而 y, z 是任意实数.

施瓦兹不等式

因为

$$2|yz| \leqslant y^2 + z^2. \tag{2.1}$$

这个不等式对于任意实数对 y, z 成立, 所以当它们的值被随机选择时不等式也成立, 即

451

$$2|YZ| \leqslant |Y|^2 + |Z|^2.$$

两边取期望, 我们得到

$$2E[|YZ|] \leqslant E[|Y|^2] + E[|Z|^2].$$

特别地, $E[|YZ|] < \infty$. 若我们在 (2.1) 中用

$$Y/\sqrt{E[|Y|^2]} = Y/\|Y\| \qquad \text{代替} y,$$
$$Z/\sqrt{E[|Z|^2]} = Z/\|Z\| \qquad \text{代替} z,$$

并取期望, 得

$$2E\left\{\frac{|YZ|}{\|Y\|\cdot\|Z\|}\right\} \leqslant E\left\{\frac{Y^2}{\|Y\|^2}\right\} + E\left\{\frac{Z^2}{\|Z\|^2}\right\} = 2.$$

因此

$$E[|YZ|] \leqslant \sqrt{E[Y^2]E[Z^2]} = \|Y\|\,\|Z\|.$$

这就是著名的**施瓦兹不等式**.

平行四边形公式

我们计算

$$\begin{aligned} E[(Y+Z)^2] &= E[Y^2 + 2YZ + Z^2] \\ &= E[Y^2] + 2E[YZ] + E[Z^2], \end{aligned}$$

和

$$\begin{aligned} E[(Y-Z)^2] &= E[Y^2 - 2YZ + Z^2] \\ &= E[Y^2] - 2E[YZ] + E[Z^2], \end{aligned}$$

把两式相加可得到平行四边形公式

$$E[(Y+Z)^2] + E[(Y-Z)^2] = 2E[Y^2] + 2E[Z^2],$$

或

$$\|Y+Z\|^2 + \|Y-Z\|^2 = 2\{\|Y\|^2 + \|Z\|^2\}. \tag{2.2}$$

这个名字源于几何中类似的等式：平行四边形两条对角线的平方和等于它的四边的平方和, 在那里 Y, Z 是向量.

三角不等式

我们计算

$$E[(Y+Z)^2]=E[Y^2]+2E[YZ]+E[Z^2].$$

由施瓦兹不等式

$$E[YZ]\leqslant E[|YZ|]\leqslant\sqrt{E[Y^2]E[Z^2]}.$$

452

因此

$$\|Y+Z\|^2\leqslant E[Y^2]+2\sqrt{E[Y^2]E[Z^2]}+E[Z^2]=(\|Y\|+\|Z\|)^2,$$

或

$$\|Y+Z\|\leqslant\|Y\|+\|Z\|. \tag{2.3}$$

其名字源于几何类似的不等式：一个三角形任意两边长度之和总是超过第三边. 只要注意到 $\|Z\|$ 相当于空间向量通常长度的概念. 三角形不等式的另一种形式有时会更加有用：

$$|(\|Y\|-\|Z\|)|\leqslant\|Y-Z\|. \tag{2.4}$$

这是由 (2.3) 得到的, 如同我们从 $|a+b|\leqslant|a|+|b|$ 推得 $|(|a|-|b|)|\leqslant|a-b|$ 一样.

m.s. 极限的唯一性

如果在三角不等式 (2.3) 中置 $Y=X_n-X'$ 和 $Z=X-X_n$, 得

$$\|X-X'\|\leqslant\|X_n-X'\|+\|X_n-X\|.$$

这样, 如果 $X_n\to X(\text{m.s.})$ 和 $X_n\to X'(\text{m.s.})$, 则 $E[(X-X')^2]=\|X-X'\|^2=0$. 利用切比雪夫不等式, 我们有

$$\Pr\{|X-X'|>\varepsilon\}\leqslant\frac{1}{\varepsilon^2}E[(X-X')^2],\quad \varepsilon>0.$$

因此对所有的正数 $\varepsilon,\Pr\{|X-X'|>\varepsilon\}=0$, 于是

$$\Pr\{X=X'\}=1. \tag{2.5}$$

因此, 随机变量序列均方收敛的极限在 (2.5) 意义下是唯一的.

依概率收敛

再次利用切比雪夫不等式

$$\Pr\{|X_n-X|>\varepsilon\}\leqslant\frac{1}{\varepsilon^2}E[(X_n-X)^2],\quad \varepsilon>0.$$

我们看到, 由 $X_n\to X(\text{m.s.})$ 推知对任意正数 ε,

$$\lim_{n\to\infty}\Pr\{|X_n-X|>\varepsilon\}=0,$$

因此, 均方收敛蕴涵着依概率收敛. 注意, 反之不一定成立!

453

二阶矩的收敛

现在我们对 $Y = X_n$ 应用三角不等式 (2.4), 有

$$\big|\,\|X\| - \|X_n\|\,\big| \leqslant \|X - X_n\| \to 0.$$

因此

$$\lim_{n\to\infty} \|X_n\| = \|X\|. \tag{2.6}$$

于是, 均方收敛可推得二阶矩收敛. 显然, 反之不一定成立.

柯西收敛准则

如果在 (2.2) 中置 $Y = X_n - X$ 和 $Z = X - X_m$, 便得到

$$E[(X_n - X_m)^2] \leqslant 2E[(X_n - X)^2] + 2E[(X_m - X)^2].$$

因此, 如果 $X_n \to X(\text{m.s.})$, 则

$$\lim_{m,n\to\infty} E[(X_n - X_m)^2] = 0$$

或

$$\lim_{m,n\to\infty} \|X_n - X_m\| = 0.$$

反之, 可以证明如果 $\{X_n\}$ 是随机变量序列且满足

$$\lim_{m,n\to\infty} E[(X_n - X_m)^2] = 0,$$

则存在一个随机变量 X 使得

$$X_n \to X(\text{m.s.}).$$

这个结果与实数列的柯西收敛准则是平行的.

均值的收敛

利用施瓦兹不等式, 令 $Z \equiv 1$, 我们有

$$E[|Y|] = E[|Y \cdot 1|] \leqslant \sqrt{E[Y^2] \cdot 1} = \|Y\|,$$

这样, 如果 $E[Y^2] < \infty$, 则有 $E[|Y|] < \infty$. 用类似方式令 $Y = X_n - X$, 我们计算

$$|E[X_n] - E[X]| \leqslant E[|X_n - X|] \leqslant \|X_n - X\|.$$

这样, 当 $n \to \infty$ 时, $X_n \to X(\text{m.s.})$ 蕴涵着

$$\lim_{n\to\infty} E[X_n] = E[X].$$

协方差的收敛

设 $\{X_n\}$ 是均方收敛于随机变量 X 的随机变量序列. 令 Y 是具有有限二阶矩的随机变量. 我们将证明 $\lim_{n\to\infty} E[YX_n] = E[YX]$, 或等价地, $\lim_{n\to\infty} |E[YX_n] - E[YX]| = 0$. 我们使用施瓦兹不等式 454

$$|E[YX_n] - E[YX]| = |E[Y(X_n - X)]| \leqslant \|Y\|\|X_n - X\|.$$

由于 $E[Y^2] < \infty$ 和 $\|X_n - X\| \to 0$. 这就完成了证明.

自回归和滑动平均过程

令 $\{X_n; n = 0, \pm1, \pm2, \cdots\}$ 是协方差平稳过程. 假设对于满足 $|\lambda| < 1$ 的某个实数 λ, 随机变量序列

$$\xi_n = X_n - \lambda X_{n-1}$$

是不相关的, 且具有零均值和共同的方差 σ^2. 这样一个过程称为 1 **阶自回归过程.** 我们有

$$\begin{aligned} X_n &= \lambda X_{n-1} + \xi_n \\ &= \lambda\{\lambda X_{n-2} + \xi_{n-1}\} + \xi_n \\ &= \lambda^2 X_{n-2} + \lambda\xi_{n-1} + \xi_n \\ &= \lambda^k X_{n-k} + \sum_{j=0}^{k-1} \lambda^j \xi_{n-j}. \end{aligned} \tag{2.7}$$

我们计算 X_n 与 $\sum_{j=0}^{k-1} \lambda^j \xi_{n-j}$ 之间的平均平方差, 得到

$$\begin{aligned} E\left[\left(X_n - \sum_{j=0}^{k-1} \lambda^j \xi_{n-j}\right)^2\right] &= E[(\lambda^k X_{n-k})^2] \\ &= \lambda^{2k} E[X_{n-k}^2]. \end{aligned}$$

由于过程是平稳的, 即 $E[X_{n-k}^2]$ 是常数, 与 n 和 k 无关, 并由于 $|\lambda| < 1$, 右边部分以几何比率递减趋于 0. 这样,

$$\begin{aligned} X_n &= \lim_{k\to\infty} \sum_{j=0}^{k-1} \lambda^j \xi_{n-j} (\text{m.s.}) \\ &= \sum_{j=0}^{\infty} \lambda^j \xi_{n-j}. \end{aligned} \tag{2.8}$$

应当记住 $\sum_{j=0}^{\infty}$ 是指部分和序列 $\sum_{j=0}^{k-1}$ 均方收敛的极限. 等式 (2.8) 把原来的过程表示

455

为滑动平均 (见本节的例 C).

由于均方收敛性蕴涵着均值的收敛和二阶矩的收敛, 我们有

$$E[X_n] = \lim_{k\to\infty} E\left[\sum_{j=0}^{k-1} \lambda^j \xi_{n-j}\right] = 0,$$

并且

$$\begin{aligned} E[X_n^2] &= \lim_{k\to\infty} E\left[\left(\sum_{j=0}^{k-1} \lambda^j \xi_{n-j}\right)^2\right] \\ &= \lim_{k\to\infty} \left\{ E\left[\sum_{j=0}^{k-1} \lambda^{2j} \xi_{n-j}^2\right] + E\left[\sum_{i\neq j} \lambda^{i+j} \xi_{n-i}\xi_{n-j}\right]\right\}. \end{aligned}$$

由于序列 $\{\xi_m\}$ 是不相关的, 第二项的期望为 0. 由 $E[\xi_{n-j}^2] = \sigma^2$ 和 $|\lambda| < 1$, 我们得到

$$\begin{aligned} E[X_n^2] &= \lim_{k\to\infty} \sum_{j=0}^{k-1} \lambda^{2j}\sigma^2 \\ &= \sigma^2/(1-|\lambda|^2). \end{aligned}$$

现计算 X_n 和 X_{n+k} 的协方差. 由 (2.7) 得

$$X_{n+k} = \lambda^k X_n + \sum_{j=0}^{k-1} \lambda^j \xi_{n+k-j},$$

因此

$$E[X_n X_{n+k}] = \lambda^k E[X_n^2] + E\left[X_n \cdot \sum_{j=0}^{k-1} \lambda^j \xi_{n+k-j}\right].$$

上式右边第一项是 $\lambda^k\sigma^2$. 利用协方差收敛, 第二项可按通常方法求值, 即

$$\begin{aligned} E\left[X_n \sum_{j=0}^{k-1} \lambda^j \xi_{n+k-j}\right] &= \lim_{l\to\infty} E\left[\sum_{m=0}^{l-1} \lambda^m \xi_{n-m} \cdot \sum_{j=0}^{k-1} \lambda^j \xi_{n+k-j}\right] \\ &= \lim_{l\to\infty} \sum_{m=0}^{l-1}\sum_{j=0}^{k-1} \lambda^{m+j} E[\xi_{n-m}\xi_{n+k-j}] \\ &= 0, \end{aligned}$$

其中我们多次利用了 $\{\xi_m\}$ 不相关这个事实. 这样

$$E[X_nX_{n+k}] = \sigma^2\lambda^k/(1-|\lambda|^2), \quad k=0,1,2,\cdots.$$

我们继续往下讨论, 把过程写成

$$X_n = \lambda X_{n-1} + \xi_n,$$

456

这提示我们协方差平稳过程的自然推广具有如下形式

$$X_n = \lambda_1 X_{n-1} + \lambda_2 X_{n-2} + \cdots + \lambda_p X_{n-p} + \xi_n, \tag{2.9}$$

其中 $\{\xi_n\}$ 是零均值不相关的随机变量序列, 具有共同的方差. 这样一个过程称为 **p 阶自回归过程.**

对于这样一个过程, 由线性回归理论可猜测当已知过去 $X_{n-1}, X_{n-2}, \cdots$ 时对于 X_n 的预测量可能由下面公式给出

$$\widehat{X}_n = \lambda_1 X_{n-1} + \cdots + \lambda_p X_{n-p}.$$

但这未必成立. 关键在于 ξ_n 和过去 $(X_{n-1},\cdots,X_{n-p})$ 的相关性. 这种预测问题是下一节要讨论的. 这里我们仅为了确定什么情况下 p 阶自回归过程具有滑动平均表示做一些准备工作.

往前一个时间单位, (2.9) 变为

$$X_{n-1} = \lambda_1 X_{n-2} + \lambda_2 X_{n-3} + \cdots + \lambda_p X_{n-p-1} + \xi_{n-1},$$

把此式代回 (2.9) 得到

$$\begin{aligned} X_n =& \lambda_1[\lambda_1 X_{n-2} + \cdots + \lambda_p X_{n-p-1} + \xi_{n-1}] \\ & + \lambda_2 X_{n-2} + \cdots + \lambda_p X_{n-p} + \xi_n \\ =& \xi_n + \lambda_1 \xi_{n-1} + (\lambda_1^2 + \lambda_2) X_{n-2} + \cdots \\ & + (\lambda_1\lambda_{p-1} + \lambda_p) X_{n-p} + \lambda_1\lambda_p X_{n-p-1}. \end{aligned}$$

按此步骤进行 m 次, 我们得到

$$\begin{aligned} X_n = & \xi_n + \delta_1 \xi_{n-1} + \cdots + \delta_m \xi_{n-m} + \beta_{m1} X_{n-m-1} \\ & + \beta_{m2} X_{n-m-2} + \cdots + \beta_{mp} X_{n-m-p}, \end{aligned} \tag{2.10}$$

这里 δ_i 和 $\beta_{m\,i}$ 是某些常数. (每次替代在右边留下 p 个连续的 X_n). 然后将

$$X_{n-m-1} = \xi_{n-m-1} + \lambda_1 X_{n-m-2} + \cdots + \lambda_p X_{n-m-p-1}$$

代入 (2.10) 得到

$$
\begin{aligned}
X_n = \xi_n + \delta_1\xi_{n-1} + \cdots + \delta_m\xi_{n-m} + \beta_{m1}\xi_{n-m-1} \\
+ (\beta_{m1}\lambda_1 + \beta_{m2})X_{n-m-2} + \cdots \\
+ (\beta_{m1}\lambda_{p-1} + \beta_{mp})X_{n-m-p} + \beta_{m1}\lambda_p X_{n-m-p-1},
\end{aligned}
$$

457

由此我们导出递归关系式

$$
\begin{aligned}
\delta_{m+1} &= \beta_{m1}, \\
\beta_{m+1,j} &= \beta_{m1}\lambda_j + \beta_{m,j+1}, \quad j = 1, \cdots, p-1,
\end{aligned}
$$

且

$$\beta_{mp} = \beta_{m1}\lambda_p.$$

把这个步骤连续进行下去, 得到滑动平均表达式

$$X_n = \sum_{k=0}^{\infty} \delta_k \xi_{n-k}, \quad \delta_0 = 1, \tag{2.11}$$

倘若当 $m \to \infty$ 时其余项 $R_m = (\beta_{m1}X_{n-m-1} + \cdots + \beta_{mp}X_{n-m-p})$ 依均方意义可忽略. 下面我们即将给出使 (2.11) 成立的适当条件.

上述过程在抽象符号下将显得更加整齐. 设 T 是位移算子, 它把实数列 $\{x_n\} = (\cdots, x_{-1}, x_0, x_1, \cdots)$ 变为序列 $T(\{x_n\}) = (\cdots, x_0, x_1, x_2, \cdots)$(这里 x_1 处于零坐标). 每个时间坐标都向前移动一个单位. 自然, T^{-1} 的功能相反:

$$T^{-1}(\{x_n\}) = (\cdots, x_{-2}, x_{-1}, x_0, \cdots),$$

且

$$T^{-k}(\{x_n\}) = (\cdots, x_{-k-1}, x_{-k}, x_{-k+1}, \cdots).$$

在这些记号下, (2.9) 变为

$$
\begin{aligned}
\{X_n\} &= \lambda_1 T^{-1}(\{X_n\}) + \cdots + \lambda_p T^{-p}(\{X_n\}) + \{\xi_n\} \\
&= \sum_{k=1}^{p} \lambda_k T^{-k}(\{X_n\}) + \{\xi_n\},
\end{aligned}
$$

或

$$\left(I - \sum_{k=1}^{p} \lambda_k T^{-k}\right)(\{X_n\}) = \{\xi_n\},$$

其中 $I = T^0$ 是恒等算子. 现在来看在什么条件下 $\left(I - \sum_{k=1}^{p} \lambda_k T^{-k}\right)$ 可逆, 使得我们有

$$\{X_n\} = \left(I - \sum_{k=1}^{p} \lambda_k T^{-k}\right)^{-1} (\{\xi_n\}).$$

即, 如果

$$X_n = \sum_{i=0}^{\infty} \delta_i \xi_{n-i} \quad \text{或} \quad \{X_n\} = \left(\sum_{i=0}^{\infty} \delta_i T^{-i}\right) (\{\xi_n\}),$$

458

则 δ_i 应该是下列展开式的系数

$$\left(1 - \sum_{k=1}^{p} \lambda_k z^k\right)^{-1} = \sum_{k=0}^{\infty} \delta_k z^k.$$

因此, 形式上可利用长除法来确定它们. 例如, 对我们先前讨论过的情况: $\lambda_1 = \lambda, \lambda_k = 0$ 对于 $k \geqslant 2$, 我们有

$$\frac{1}{1 - \lambda z} = 1 + \lambda z + \lambda^2 z^2 + \lambda^3 z^3 + \cdots,$$

或

$$X_n = \sum_{j=0}^{\infty} \lambda^j \xi_{n-j},$$

这个已由 (2.8) 得到.

经过冗长的直接计算我们有

$$\frac{1}{1 - \lambda_1 z - \cdots - \lambda_p z^p} = 1 + \delta_1 z + \cdots + \delta_m z^m + r_m(z),$$

其中

$$r_m(z) = \frac{\beta_{m1} z^{m+1} + \cdots + \beta_{mp} z^{m+p}}{1 - \lambda_1 z - \cdots - \lambda_p z^p}.$$

即, 形式上的除法正好得到我们前面通过直接代入法所得到的相同结果. 由此推得, 如果当 $m \to \infty$ 时 $r_m(z) \to 0$, 则每个 $\beta_{mk} \to 0, k = 1, \cdots, p$, 并且随机余项 $R_m = \beta_{m1} X_{n-m-1} + \cdots + \beta_{mp} X_{n-m-p}$ 均方趋于 0, 所以无限滑动平均表达式 (2.11) 是正确的. 我们下面给出条件. 方程

$$x^p - \lambda_1 x^{p-1} - \cdots - \lambda_p = 0 \tag{2.12}$$

有 p 个根: $x_1, \cdots, x_p$. 如果 $|x_i| < 1, i = 1, \cdots, p$, 则

$$1 - \lambda_1 z - \cdots - \lambda_p z^p = 0$$

的根是 $z_i = 1/x_i$, 倘若 $\lambda_p \neq 0$ 且 $|z_i| > 1$. 对于任何满足 $|z| < \min|z_i|$ 的 z, 级数

$$\frac{1}{1-\lambda_1 z - \cdots - \lambda_p z^p} = \frac{1}{\prod_{i=1}^{p}\left(1-\frac{z}{z_i}\right)}$$
$$= \prod_{i=1}^{p}\sum_{v=0}^{\infty}\left(\frac{z}{z_i}\right)^v = \sum_{r=0}^{\infty}\delta_r z^r,$$

绝对收敛. 因此, 当 $m \to \infty$ 时对于 $|z| < \min_i |z_i|$(特别地, 当 $|z| = 1$ 时), $r_m(z) \to 0$. 由此推得当 $m \to \infty$ 时对于 $k = 1, 2, \cdots, p, \beta_{mk} \to 0$. 这时 R_m 均方收敛于 0. 所以无限滑动平均表达式 (2.11) 成立.

为了以后参考用, 基于前面分析, 我们强调一个重要结论:

附注 2.1 假定 $\{X_n\}$ 是零均值协方差平稳过程且满足

$$X_n = \lambda_1 X_{n-1} + \cdots + \lambda_p X_{n-p} + \xi_n, \quad n = 0, \pm 1, \cdots,$$

其中 $\{\xi_n\}$ 是零均值不相关随机变量序列, 且 $\lambda_1, \cdots, \lambda_p$ 是固定的. 如果

$$x^p - \lambda_1 x^{p-1} - \cdots - \lambda_p = 0$$

的每个根 $x_i, i = 1, 2, \cdots, p$, 有 $|x_i| < 1$, 则 ξ_n 和 $(X_{n-1}, \cdots, X_{n-p})$ 是不相关的. 这是由于 $X_{n-k} = \sum_{j=0}^{\infty}\delta_j \xi_{n-k-j} (k \geqslant 1)$ 中的各项与 ξ_n 依假设是不相关的.

为了了解含有相反方向滑动平均的另一种情况, 我们看一看

$$X_n = \eta_n + \rho\eta_{n+1} + \rho^2\eta_{n+2} + \cdots, \tag{2.13}$$

其中, $|\rho| < 1$ 且 $\{\eta_n\}$ 是零均值不相关的. 那么

$$\rho X_n = X_{n-1} - \eta_{n-1}, \quad \text{或} \quad X_n = \lambda X_{n-1} + \xi_n,$$

其中, $\lambda = 1/\rho, \xi_n = -\rho^{-1}\eta_{n-1}$. 由于 $\{\xi_n\}$ 是不相关的, 所以 $\{X_n\}$ 是自回归过程, 但 $|\lambda| > 1$ 排除了形如 (2.11) 的滑动平均表达式. 在 (2.13) 中过程的构造是向前方向的滑动平均, 当 (2.12) 的根按绝对值全部超过 1 时, 呈现出典型性质. 当 (2.12) 的根的绝对值位于 1 的两侧, 为表示 $\{X_n\}$ 要使用双无限滑动平均 $\sum_{k=-\infty}^{+\infty}\delta_k\xi_{n-k}$. 最后, 如果每个根的绝对值都正好为 1, 则满足 (2.9) 的自回归过程将不再是平稳的 (除了平凡情形 $\xi_n \equiv 0$).

自回归和滑动平均模型可以结合起来产生更加复杂的平稳过程. 我们可以假设

$$X_n = \lambda_1 X_{n-1} + \cdots + \lambda_p X_{n-p} + \eta_n,$$

其中

$$\eta_n = \alpha_1 \xi_n + \alpha_2 \xi_{n-1} + \cdots + \alpha_m \xi_{n-m+1},$$

460

是一个滑动平均. 更加一般地, 我们可以考虑

$$Z_n = X_n + \varepsilon_n,$$

其中 $\{\varepsilon_n\}$ 是一个不相关的“噪声”过程. 这些比较复杂模型的分析是十分困难的.

9.3 平均平方误差预测

出现在很多领域中的一个问题是关于预测将要观察的给定随机变量值的问题. 报纸上常常报导关于将来国民生产总值、就业人数、日用品价格和其他经济量的预测值. 控制问题, 例如过程控制, 也隐含着涉及到预测. 如果一个过程的预测结果不令人满意, 可通过控制或改革使过程与我们的希望更加接近.

本节讨论关于平稳随机过程的预测问题. 假设在时刻 $n = \cdots, -2, -1, 0, 1, 2, \cdots$ 我们所关心的量构成一个平稳过程 $\{X_n\}$. 现在要根据已观测的值 $X_n, X_{n-1}, X_{n-2}, \cdots$ 来预测 X_{n+1}, 或者更加一般地预测 X_{n+k}. 这就是平稳过程的**预测问题**或**外推问题**.

一般预测理论

设 X 表示某个“试验”的结果. 假设 X 是未知的, 是将要观测的, 现在希望预测将要出现的 X 的值. 设 $\widehat{X}$ 是 X 的预测值. 预测误差为 $X - \widehat{X}$, 我们希望这个误差在某种意义下尽可能的小. 这时马上会出现一个问题:“小”是什么意思? 一般地说, $X - \widehat{X}$ 是一个随机变量, 因此我们所说的小是**按平均意义**, 即误差的某个函数的期望值应当是小的. 究竟应选择误差的什么函数呢? 比较合理的要求是此函数当误差为 0 时取最小值, 随着误差绝对值的增大而增加. 在本节, 我们利用平均平方误差 $E[(X - \widehat{X})^2] = \|X - \widehat{X}\|^2$ 作为衡量标准. 这个标准体现了大误差比小误差影响强烈这个合理的要求. 然而更重要的是, 这个标准在数学上便于处理, 我们将会看到, 它能使一般预测理论得以发展, 例如, 在若干特定情况下能导出最佳预测直接表达式.

461

这样, 我们的问题是在所有预测量 $\widehat{X}$ 中寻找使平均平方误差 $E[(X - \widehat{X})^2]$ 最小的预测量. 为了使得描述完备, 我们必须给定所允许的预测量类, 以便在该类中寻求最佳的预测量. 一般来说, 这个类是由所涉及的试验或 X 的分布所提供的信息来确定的. 下面我们用 $\mathbf{H}$ 来表示所有允许的预测量类.

例　(i) 令 X 为试验结果, 已知其均值为 μ, 方差为 σ^2. 在均方意义下并且在缺乏更多信息的情况下, X 的最佳预测是 $\widehat{X}=\mu$. 由于对 X 没有更多的信息可用, 我们取预测量空间 $\mathbf{H}$ 为实数集合. 即可以选择任何实数 a 作为 X 的预测值. 我们计算

$$
\begin{aligned}
&E[(X-\widehat{X})^2]=E[(X-\mu+\mu-\widehat{X})^2]\\
&=E[(X-\mu)^2]+2E[(X-\mu)(\mu-\widehat{X})]+E[(\mu-\widehat{X})^2].
\end{aligned}
$$

已知, $E[(X-\mu)^2]=\sigma^2$, 并且由于 $\widehat{X}$ 是固定实数, 所以 $E[(\mu-\widehat{X})^2]=(\mu-\widehat{X})^2$ 且 $E[(X-\mu)(\mu-\widehat{X})]=(\mu-\widehat{X})E[X-\mu]=0$. 从而,

$$
E[(X-\widehat{X})^2]=\sigma^2+(\mu-\widehat{X})^2.
$$

由此立刻推得, 当置 $\widehat{X}=\mu$ 时, 均方误差最小. 这样, 在没有更多信息的情况下, 随机变量 X 最小均方误差预测量是它的均值 $\mu=E[X]$.

(ii) 现在令 X,Y 是具有有限方差和已知联合分布的随机变量. 我们假设 X 是由对 Y 的一次观测来预测的. 例如, 钢的拉力强度的预测可通过观测它的硬度; 物理量如压力、温度等等, 其实际值可根据带有随机误差测量值来预测.

由于我们假设 Y 是可观测的, 并且关于 X 的结果没有更多信息可用 (除了知道其联合分布之外), 我们允许任何具有有限方差的函数 $\widehat{X}=f(Y)$ 作为其预测量.

这时可能会猜想, 在观察值 $Y=y$ 之后, 我们处于与例 (i) 相同的情况, 只是在第二种情况下相应的分布是已知 $Y=y$ 情况下 X 的条件分布. 因此最小均方误差预测量应该是 X 关于条件分布的均值 $\mu_{X|Y}=E[X|Y]$. 事实确实如此. 我们计算

$$
\begin{aligned}
&E[(X-\widehat{X})^2]\\
&=E[(X-\mu_{X|Y})^2]+2E[(X-\mu_{X|Y})(\mu_{X|Y}-\widehat{X})]\\
&\quad+E[(\mu_{X|Y}-\widehat{X})^2].
\end{aligned}
$$

462

我们证明在右边的中间项的期望为 0. 注意到 $\mu_{X|Y}$ 和 $\widehat{X}$ 是 Y 的函数, 由于 $E[(X-\mu_{X|Y})|Y]=E[X|Y]-\mu_{X|Y}=0$. 我们有

$$
\begin{aligned}
&E[(X-\mu_{X|Y})(\mu_{X|Y}-\widehat{X})]\\
&=E\{E[(X-\mu_{X|Y})(\mu_{X|Y}-\widehat{X})|Y]\}\\
&=E\{(\mu_{X|Y}-\widehat{X})E[(X-\mu_{X|Y})|Y]\}\\
&=0,
\end{aligned}
$$

这样,

$$
E[(X-\widehat{X})^2]=E[(X-\mu_{X|Y})^2]+E[(\mu_{X|Y}-\widehat{X})^2],
$$

右边部分当 $\widehat{X} = E[X|Y]$ 时取得最小值.

(iii) 为了计算 $\widehat{X} = E[X|Y]$, 我们需要知道 X, Y 联合分布的全部信息. 即使这些信息可以得到, 导出公式常常也是太复杂而不便于实际求值. 在研究协方差平稳过程中, 我们假定仅知道前两阶矩的信息, 而不提供联合分布其他进一步的信息. 因此希望有一个预测理论, 既能得到简单的预测公式, 又只需要前两阶矩的信息. 这就是所谓的**线性预测**. 设 X、Y 分别具有均值 μ_X、μ_Y, 方差 σ_X^2、σ_Y^2 以及协方差 $\sigma_{X,Y} = E[(X-\mu_X)(Y-\mu_Y)]$. 我们要求用 Y 的线性函数作为 X 的预测量, 例如 $\widehat{X} = a + bY$, 或等价地, $\widehat{X} = a + b(Y - \mu_Y)$, 其中 a 和 b 是任意实数. 由于本例的允许预测类比例 (ii) 中的小, 因此所导出的最小均方预测的误差不会比例 (ii) 中的小, 很可能会比较大. 然而预测公式是线性的, 比较简单, 并且下面我们将指出, 最佳系数由已知的矩就能确定, 不必知道整个联合分布. 下面我们来证明, 用 Y 预测 X 的最佳线性预测公式是

$$\widehat{X}^* = \mu_X + \frac{\sigma_{XY}}{\sigma_Y^2}(Y - \mu_Y).$$

令

$$\widehat{X} = a + b(Y - \mu_Y),$$

且注意到,

$$\widehat{X}^* - \widehat{X} = a' + b'(Y - \mu_Y),$$

463

其中, $a' = \mu_X - a, b' = (\sigma_{XY}/\sigma_Y^2) - b$. 同前面一样, 计算

$$\begin{aligned}
&E[(X-\widehat{X})^2] \\
&= E[(X-\widehat{X}^*)^2] + 2E[(X-\widehat{X}^*)(\widehat{X}^*-\widehat{X})] + E[(\widehat{X}^*-\widehat{X})^2].
\end{aligned}$$

我们证明右边第二项为 0. 事实上,

$$\begin{aligned}
&E[(X-\widehat{X}^*)(\widehat{X}^*-\widehat{X})] \\
&= E\left[\left\{(X-\mu_X) - \frac{\sigma_{XY}}{\sigma_Y^2}(Y-\mu_Y)\right\} \times \{a' + b'(Y-\mu_Y)\}\right] \\
&= a'E\left[(X-\mu_X) - \frac{\sigma_{XY}}{\sigma_Y^2}(Y-\mu_Y)\right] \\
&\quad + b'E\left[(X-\mu_X)(Y-\mu_Y) - \frac{\sigma_{XY}}{\sigma_Y^2}(Y-\mu_Y)^2\right] \\
&= 0 + b'\left(\sigma_{XY} - \frac{\sigma_{XY}}{\sigma_Y^2}\sigma_Y^2\right) \\
&= 0.
\end{aligned}$$

故

$$E[(X-\widehat{X})^2] = E[(X-\widehat{X}^*)^2] + E[(\widehat{X}^*-\widehat{X})^2],$$

并且右边部分当 $\widehat{X}=\widehat{X}^*$ 时最小. 这样, $\widehat{X}^*$ 就是所求的最小均方误差线性预测量.

有一个重要的特殊情况, 最小均方误差的线性预测量在线性的限制下并没有降低预测效果. 如果 X 和 Y 具有联合正态分布, 则如第 1 章所指出的那样, 已知 $Y=y, X$ 的条件均值是

$$E[X|Y=y]=\mu_X+\frac{\sigma_{XY}}{\sigma_Y^2}(y-\mu_Y).$$

我们看到这个最佳预测量是 y 的线性函数. 这样, 如果 X 和 Y 具有联合正态分布, 在已知 Y 的条件下, X 的最小均方误差线性预测量事实上就是最小均方误差预测量.

上面这些例子给出了一种模式: 在每种情况下, 为了说明给定的预测量 $\widehat{X}^*$ 是最佳的, 问题的关键都在于证明叉积 $E[(X-\widehat{X}^*)(\widehat{X}^*-\widehat{X})]$ 为零. 我们可将此归结为下面的最小均方误差预测定理:

464

定理 3.1 设 X 是具有有限二阶矩的随机变量, 令 $\mathbf{H}$ 是允许预测量 $\widehat{X}$ 的空间, 即某个具有有限二阶矩随机变量 $\widehat{X}$ 的集合. 假设 $\mathbf{H}$ 是如下意义下的线性空间: 当 $\widehat{X}_1$ 和 $\widehat{X}_2$ 是预测量而 a 是实数时, 则 $a\widehat{X}_1+\widehat{X}_2$ 也可作为预测量. 那么:

(i) 预测量 $\widehat{X}^*$ 具有最小均方误差当且仅当对于每个预测量 $U, E[(X-\widehat{X}^*)U]=0$;

(ii) 如果最小均方误差预测量存在, 它在均方距离意义下是唯一的, 即, 如果 $\widehat{X}_1^*$ 和 $\widehat{X}_2^*$ 是最小均方误差预测量, 则 $E[(\widehat{X}_1^*-\widehat{X}_2^*)^2]=0$;

(iii) 最小均方误差预测量存在的充分条件是, $\mathbf{H}$ 在如下意义下是闭的: 如果 $\widehat{X}_n, n=0,1,\cdots$ 是一列均方收敛于随机变量 $\widehat{X}$ 的预测量, 则 $\widehat{X}$ 也是预测量.

证明 (i) 假设 $\widehat{X}^*$ 是一个预测量, $E[(X-\widehat{X}^*)\cdot U]=0$ 对每个预测量 U 成立. 我们来证明 $\widehat{X}^*$ 是具有最小均方误差的预测量. 令 $\widehat{X}$ 是任意一个允许的预测量, 计算

$$E[(X-\widehat{X})^2]=E[(X-\widehat{X}^*)^2]+2E[(X-\widehat{X}^*)(\widehat{X}^*-\widehat{X})]+E[(\widehat{X}^*-\widehat{X})^2].$$

置 $U=\widehat{X}^*-\widehat{X}$. 则 U 也是一个预测量, 这是因为 $\widehat{X}^*$ 和 $\widehat{X}$ 都是预测量, 其任意线性组合还是一个允许的预测量. 这样,

$$E[(X-\widehat{X}^*)(\widehat{X}^*-\widehat{X})]=E[(X-\widehat{X}^*)U]=0,$$

且

$$E[(X-\widehat{X})^2]=E[(X-\widehat{X}^*)^2]+E[(\widehat{X}^*-\widehat{X})^2].$$

因此,

$$E[(X-\widehat{X})^2]\geqslant E[(X-\widehat{X}^*)^2],$$

即 $\widehat{X}^*$ 的均方误差比任何其他预测量 $\widehat{X}$ 都来得小. 故 $\widehat{X}^*$ 是最小均方误差预测量.

另一方面, 给定预测量 $\widehat{X}^*$, 假设 $E[(X-\widehat{X}^*)\cdot U]=a\neq 0$ 对某个预测量 U 成立. 我们将证明

$$\widehat{X}=\widehat{X}^*+\frac{a}{E[U^2]}U \tag{3.1}$$

具有较小的均方误差, 因此 $\widehat{X}^*$ 不是最佳的, 这就完成了 (i) 的证明. 首先注意到 $\widehat{X}$ 是预测量的线性组合, 因而也是一个预测量. 改写 (3.1) 为

465

$$\widehat{X}-\widehat{X}^*=\frac{a}{E[U^2]}U,$$

我们有

$$\begin{aligned}E[(X-\widehat{X})^2]&=E[\{(X-\widehat{X}^*)-(\widehat{X}-\widehat{X}^*)\}^2]\\&=E[(X-\widehat{X}^*)^2]-2E[(X-\widehat{X}^*)(\widehat{X}-\widehat{X}^*)]+E[(\widehat{X}-\widehat{X}^*)^2],\end{aligned}$$

于是得到

$$\begin{aligned}&E[(X-\widehat{X})^2]\\&=E[(X-\widehat{X}^*)^2]-2\frac{a}{E[U^2]}E[(X-\widehat{X}^*)U]+\frac{a^2}{\{E[U^2]\}^2}E[U^2]\\&=E[(X-\widehat{X}^*)^2]-\frac{a^2}{E[U^2]}\\&<E[(X-\widehat{X}^*)^2].\end{aligned}$$

即我们证明了 $\widehat{X}$ 具有较小的均方误差.

(ii) 为了证明最小均方误差预测量是唯一的, 我们假设 $\widehat{X}_1^*$ 和 X_2^* 都具有最小均方误差. 则对于所有预测量 U 有 $E[(X-\widehat{X}_i^*)U]=E[XU-\widehat{X}_i^*U]=E[XU]-E[\widehat{X}_i^*U]=0, i=1,2$. 这样,

$$E[\widehat{X}_1^*U]=E[XU]=E[\widehat{X}_2^*U],$$

及

$$E[(\widehat{X}_1^*-\widehat{X}_2^*)U]=0,$$

对于所有的预测量 U 都成立. 我们选择 $U=\widehat{X}_1^*-\widehat{X}_2^*$, 得到

$$E[(\widehat{X}_1^*-\widehat{X}_2^*)^2]=0. \tag{3.2}$$

故, 最小均方误差预测量在均方距离的意义下是唯一的. 利用切比雪夫不等式可以证明, 如本章前面所做一样, 由 (3.2) 可得

466

$$\Pr\{\widehat{X}_1^*\neq\widehat{X}_2^*\}=0.$$

(iii) 令 d 为 $\mathbf{H}$ 中预测量均方误差的下确界, 即

$$d = \inf\{E[|X - \widehat{X}|^2] : \widehat{X} \in \mathbf{H}\}.$$

在 $\mathbf{H}$ 中我们选择一列预测量 $\{\widehat{X}_n\}$, 使其均方误差趋于 d,

$$\lim_{n\to\infty} E[|X - \widehat{X}_n|^2] = d.$$

对任意的 m, n, 我们应用平行四边形公式于

$$Y = X - \widehat{X}_n, \quad Z = X - \widehat{X}_m,$$

得到

$$\begin{aligned} &E[|\widehat{X}_m - \widehat{X}_n|^2] + E[|2X - \widehat{X}_n - \widehat{X}_m|^2] \\ &= 2E[|X - \widehat{X}_n|^2] + 2E[|X - \widehat{X}_m|^2]. \end{aligned} \tag{3.3}$$

由于 $\frac{1}{2}(\widehat{X}_n + \widehat{X}_m)$ 是预测量, 它的均方误差必超过 d. 因此,

$$E[|2X - \widehat{X}_n - \widehat{X}_m|^2] = 4E[|X - \frac{1}{2}(\widehat{X}_n + \widehat{X}_m)|^2] \geqslant 4d.$$

将此式代入 (3.3) 得到

$$E[|\widehat{X}_n - \widehat{X}_m|^2] + 4d \leqslant 2E[|\widehat{X}_n - X|^2] + 2E[|\widehat{X}_m - X|^2].$$

上式两边同时减去 $4d$, 并令 m, n 无限增加, 推得

$$\begin{aligned} &\lim_{m,n\to\infty} E[|\widehat{X}_m - \widehat{X}_n|^2] \\ &\leqslant 2\lim_{n\to\infty} E[|\widehat{X}_n - X|^2] + 2\lim_{m\to\infty} E[|\widehat{X}_m - X|^2] - 4d \\ &= 0. \end{aligned}$$

由柯西均方收敛准则可知, 存在一个随机变量, 记为 $\widehat{X}^*$, 使得 $\widehat{X}_n \to \widehat{X}^*(m.s.)$. 由假设, $\widehat{X}^*$ 是一个预测量. 这样 $X - \widehat{X}_n \to X - \widehat{X}^*(m.s.)$, 并由均方收敛下均方距离的连续性, 我们有

$$E[|X - \widehat{X}^*|^2] = \lim_{n\to\infty} E[|X - \widehat{X}_n|^2] = d.$$

故, $\widehat{X}^*$ 是最小均方误差预测量, 这就完成了定理的证明. ■

在例 (i)~(iii) 中, 我们的工作是验证已知的预测量是最佳的. 现在我们说明如何藉助于上述定理导出最佳预测量. 我们使用定理的第 (i) 部分, 即预测量具有最小均方误差的充分必要条件. 在例 (i) 中, **H** 是实数的集合, 我们希望某实数 $\widehat{X}^*$ 对于所有的实数 u 满足

467

$$E[(X-\widehat{X}^*)u]=0, \tag{3.4}$$

我们只需要考虑 $u=1$ 的情形, 此时

$$E[(X-\widehat{X}^*)]=0,$$

由于 $\widehat{X}^*$ 是非随机的, 所以

$$\widehat{X}^*=E[\widehat{X}^*]=E[X]=\mu.$$

而这样选择的 $\widehat{X}^*$ 满足式 (3.4), 由此可知 $\widehat{X}^*$ 是 (唯一) 最小均方误差预测量.

对于例 (ii), **H** 是 Y 的所有具有有限二阶矩的函数 $\widehat{X}=f(Y)$ 的集合. $\widehat{X}^*$(它是 Y 的一个函数) 为最佳的充分必要条件是对于所有的具有有限二阶矩函数 $f(Y)$,

$$E[(X-\widehat{X}^*)f(Y)]=0,$$

或

$$E[Xf(Y)]=E[\widehat{X}^*f(Y)].$$

由此推得 $\widehat{X}^*$ 是已知 Y 的条件下 X 的条件期望 (见第 1 章). 于是, $\widehat{X}^*=E[X|Y]$ 是 (在均方距离意义下唯一) 最佳预测量.

在例 (iii), 条件是对于任何 Y 的线性函数 U,

$$E[(X-\widehat{X}^*)U]=0, \tag{3.5}$$

我们记 U 和 $\widehat{X}^*$ 的线性表示式如下:

$$\begin{aligned} U &= a+b(Y-\mu_Y), \\ \widehat{X}^* &= a^*+b^*(Y-\mu_Y), \end{aligned}$$

所以式 (3.5) 变成

$$E[\{X-(a^*+b^*(Y-\mu_Y))\}\times\{a+b(Y-\mu_Y)\}]=0.$$

这式子必须对于所有的实数 a,b 都成立. 特别地, 我们先取 $a=1$ 和 $b=0$, 然后取 $a=0$ 和 $b=1$, 得到

$$E[X-(a^*+b^*(Y-\mu_Y))]=0,$$

468

或

$$E[X] = a^* - b^*E[Y - \mu_Y] = a^*,$$

以及

$$E[\{X - a^* - b^*(Y - \mu_Y)\}(Y - \mu_Y)] = 0,$$

或

$$E[(X - \mu_X)(Y - \mu_Y)] = b^*E[(Y - \mu_Y)^2],$$
$$\sigma_{XY} = b^*\sigma_Y^2.$$

这样, $a^* = \mu_X$ 和 $b^* = \sigma_{XY}/\sigma_Y^2$, 所以

$$\widehat{X}^* = \mu_X + \frac{\sigma_{XY}}{\sigma_Y^2}(Y - \mu_Y).$$

注记 审视有限维欧几里得几何中某些类似结果有助于我们更好地理解在均方误差度量下为什么能够发展漂亮的预测理论. 我们记得, 一个 (实)Bold 体是某个以向量为元素的集合, 在其中定义了向量加法运算和数乘运算, 当 x 与 y 是该集合的向量而 a 和 b 是实数时, $ax+by$ 仍是该集合的向量. 在向量空间中 Bold 体是一对向量 x, y 的实值函数, 用 (x, y) 来表示, 满足 $(x,y) = (y,x)$, $(a_1x_1+a_2x_2, y) = a_1(x_1, y)+ a_2(x_2, y)$, 以及若 x 不是零向量 $(x,x) > 0$. 在一个由积空间可以定义向量 x 的**范数**或长度为 $(x,x)^{1/2}$, 记为 $\|x\|$. 一个 Bold 体是一个完备的内积空间, 完备性是在范数所确定的度量意义下. 这意味着如果 $\{x_n\}$ 是一列向量且满足 $\lim\limits_{m,n\to\infty} \|x_n - x_m\| = 0$, 则存在一个向量 x, 使得 $\lim\limits_{n\to\infty} \|x_n - x\| = 0$.

在定理 3.1 的假设下, 预测量空间 $\mathbf{H}$ 是一个希尔伯特空间, 其内积是 $(\widehat{X}, \widehat{Y}) = E[\widehat{X}\widehat{Y}]$. 空间 $\mathbf{H}'$ 由所有形如 $aX + b\widehat{X}$ 的随机变量组成, 其中 a 和 b 是实的, $\widehat{X}$ 属于 $\mathbf{H}$, 即是预测量, 则 $\mathbf{H}'$ 也是一个希尔伯特空间. 均方误差是向量差 $X - \widehat{X}$ 的范数或长度的平方, 即它是 X 和 $\widehat{X}$ 之间距离的平方. 用希尔伯特空间术语, 我们的问题是在 $\mathbf{H}$ 中寻求一个向量 $\widehat{X}^*$, 使其最靠近 $\mathbf{H}'$ 中的向量 X.

希尔伯特空间是 d 维欧几里得空间的推广, 欧氏向量 $x = (x_1, \cdots, x_d), y = (y_1, \cdots, y_d)$, 其内积为 $(x,y) = \sum x_iy_i$, 范数为欧几里得距离 $\|x\| = \left\{\sum x_i^2\right\}^{1/2}$. 两个向量 x, y 夹角的余弦通常定义为 $(x,y)/(\|x\|\|y\|)$. 特别地, 两个向量 x, y 垂直当且仅当 $(x,y) = 0$.

问题是在空间 $\mathbf{H}'$ 的子空间 $\mathbf{H}$ 中求得一个向量 $\widehat{X}^*$, 使得它最靠近 $\mathbf{H}'$ 的向量 X. 定理 3.1 指出, $\widehat{X}^*$ 最靠近 X 当且仅当差 $X - \widehat{X}^*$ 垂直于 $\mathbf{H}$ 中所有向量 u. 图 9-1 帮助我们把问题几何直观化.

469

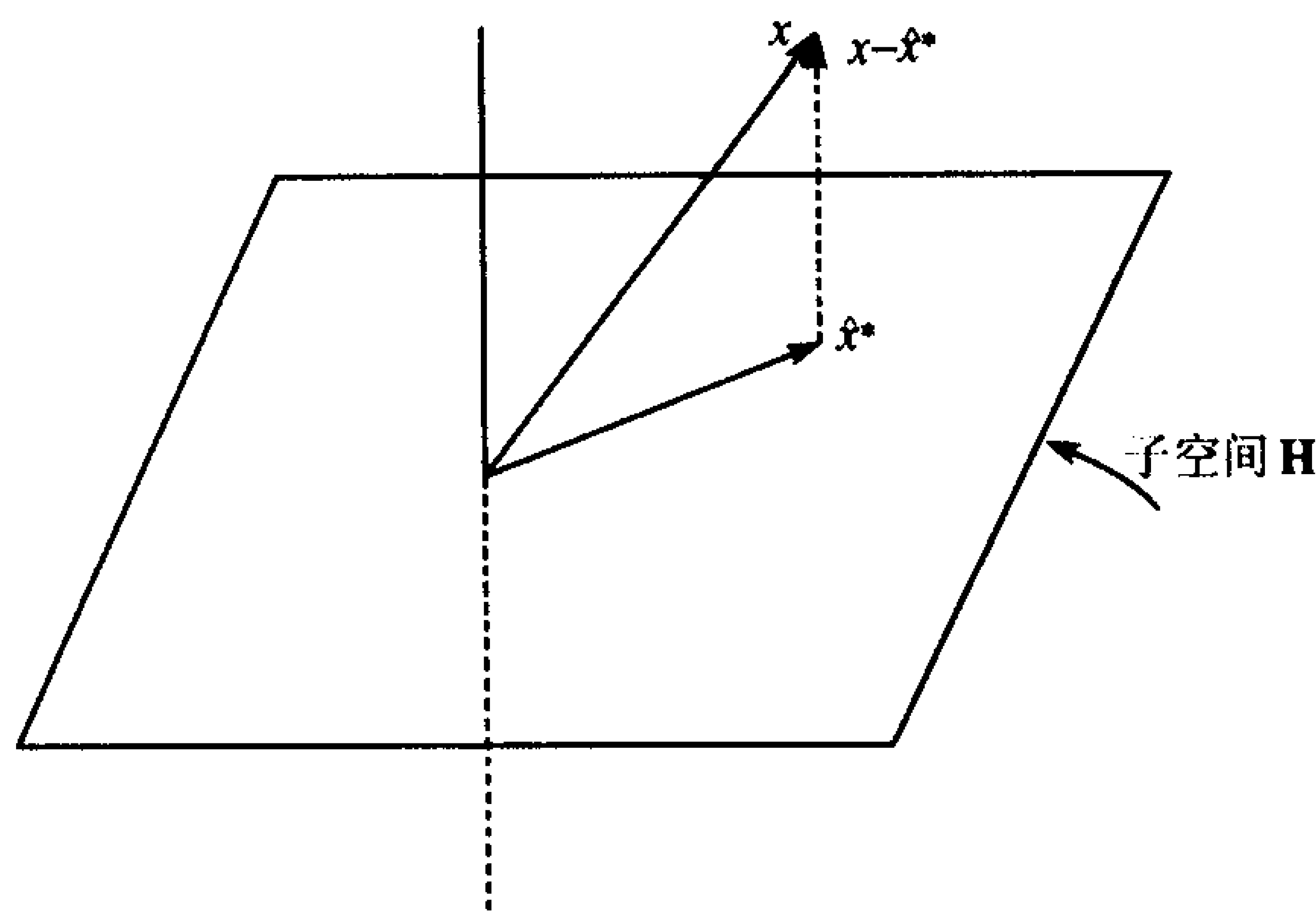


图 9-1 图形说明在子空间 **H** 中的向量 $\widehat{x}^*$ 最靠近向量 x 当且仅当差 $x-\widehat{x}^*$ 垂直于 **H** 中所有的量.

9.4 协方差平稳过程的预测

设 $X_n, n=0,\pm1,\pm2,\cdots$ 是协方差平稳过程, 它具有已知协方差函数 $R(u), u=0,\pm1,\cdots$.(其中 X_n 是实值过程, 因而 $R(-u)=R(u)$.) 过程的均值是常数, 假设是已知的. 这样, 如果需要, 可以改变原点, 我们不妨假设均值为 0. 现在我们来考虑这样一个问题: 如何根据有限过去值 $X_{n-1}, X_{n-2}, \cdots, X_{n-p}$ 来求得 X_n 的最小均方误差线性预测量? 设此预测量具有形式

$$\widehat{X}_n = \alpha_1 X_{n-1} + \alpha_2 X_{n-2} + \cdots + \alpha_p X_{n-p}.$$

利用定理 3.1 可知

$$\widehat{X}_n^* = \alpha_1^* X_{n-1} + \alpha_2^* X_{n-2} + \cdots + \alpha_p^* X_{n-p}$$

具有最小均方误差, 当且仅当对于所有的预测量

$$U = u_1 X_{n-1} + u_2 X_{n-2} + \cdots + u_p X_{n-p},$$

我们有

$$E[(X_n - \widehat{X}_n^*)U] = 0.$$

现在我们来考虑关于 U 的 p 个特别选取:

$$U_i = X_{n-i}, \quad i = 1, \cdots, p.$$

由于任何预测量 U 可以写为 $U = u_1U_1 + \cdots + u_pU_p$, 因此

470

$$E[(X_n - \widehat{X}_n^*)U] = 0 \quad \text{对于所有的 } U \text{ 成立},$$

当且仅当

$$E[(X_n - \widehat{X}_n^*)U_i] = 0, \quad i = 1, \cdots, p. \tag{4.1}$$

将 $\widehat{X}^*$ 的表示式代入 (4.1), 得

$$E[(X_n - \{\alpha_1^* X_{n-1} + \alpha_2^* X_{n-2} + \cdots + \alpha_p^* X_{n-p}\})X_{n-i}] = 0,$$

或

$$E[X_n X_{n-i}] = \alpha_1^* E[X_{n-1}X_{n-i}] + \cdots + \alpha_p^* E[X_{n-p}X_{n-i}].$$

由于这些叉积是协方差, 所以我们有

$$R(i) = \alpha_1^* R(i-1) + \cdots + \alpha_p^* R(i-p), \quad i = 1, \cdots, p.$$

这样 $\alpha^* = (\alpha_1^*, \cdots, \alpha_p^*)$ 可以取为满足 p 个线性方程的任何向量

$$\begin{aligned}
R(1) &= \alpha_1^* R(0) + \alpha_2^* R(1) + \alpha_3^* R(2) + \cdots + \alpha_p^* R(p-1), \\
R(2) &= \alpha_1^* R(1) + \alpha_2^* R(0) + \alpha_3^* R(1) + \cdots + \alpha_p^* R(p-2), \\
R(3) &= \alpha_1^* R(2) + \alpha_2^* R(1) + \alpha_3^* R(0) + \cdots + \alpha_p^* R(p-3), \\
&\vdots \\
R(p) &= \alpha_1^* R(p-1) + \alpha_2^* R(p-2) + \alpha_3^* R(p-3) + \cdots + \alpha_p^* R(0).
\end{aligned} \tag{4.2}$$

定理 3.1 的 (iii) 断言在这种情况下存在一个最小均方误差预测量. (可以证明定理所要求的假设是满足的.) 这样, (4.2) 至少存在一个解 α^*, 或许可能很多. 然而由定理 3.1 的唯一性部分知, 所有的解都导出相同的 $\widehat{X}^*$. 作为例子, 考虑 9.1 节中极端过程 $\{Z_n\}$, 其中 $Z_n = Z, n = 0, \pm1, \cdots$, 并且对于所有的 $u = 0, \pm1, \cdots, R(u) = \sigma^2$. 那么, 任何向量 α^* 若满足 $\alpha_1^* + \cdots + \alpha_p^* = 1$, 则一定是 (4.2) 的解. 然而所有这样的向量 α^* 都导致 $\widehat{Z}_n^* = Z$.

我们计算最小均方误差 σ_p^2 如下:

$$\begin{aligned}
\sigma_p^2 &= E[|X_n - \widehat{X}_n^*|^2] \\
&= E[(X_n - \widehat{X}_n^*)(X_n - \widehat{X}_n^*)] \\
&= E[(X_n - \widehat{X}_n^*)X_n] - E[(X_n - \widehat{X}_n^*)\widehat{X}_n^*].
\end{aligned}$$

第二项为 0, 这是由最佳预测量的充分必要条件推得的. 因此

$$\begin{aligned}\sigma_p^2 &= E[X_n^2] - E[\widehat{X}_n^* X_n] \\ &= R(0) - E\left[\sum_{k=1}^{p} \alpha_k^* X_{n-k} X_n\right] \\ &= R(0) - \sum_{k=1}^{p} \alpha_k^* R(k).\end{aligned}$$

471

现在考虑根据全部过去 $X_{n-1}, X_{n-2}, \cdots$ 对 X_n 进行线性预测的问题. 我们把所有具有如下形式的随机变量

$$\widehat{X}_n = \alpha_1 X_{n-1} + \cdots + \alpha_p X_{n-p},$$

作为预测量, 其中 p 是任意正整数, $\alpha_1, \alpha_2, \cdots, \alpha_p$ 是实数, 并且所有这样形式的随机变量序列在均方意义下的极限也作为预测量. 这样的预测量空间 **H** 满足基本预测定理 3.1 的所有条件. 显然, 形如

$$\widehat{X}_n = \sum_{k=1}^{\infty} \alpha_k X_{n-k} \tag{4.3}$$

的每个随机变量是一个预测量, 倘若选择 α_k 使得此无限和依均方意义收敛. 利用柯西准则于部分和序列, 可以证明只要 $\sum_{k,l} \alpha_k \alpha_l R(|k-l|) < \infty$, 无限和一定收敛. 然而, 并不是每一个有限预测量的极限都能表示为 (4.3) 的形式, 这似乎与直观不符合, 可它却是事实.

例子是容易举的. 设 $Z, \cdots \zeta_{-1}, \zeta_0, \zeta_1, \cdots$ 是一列具有零均值. 单位方差的独立同分布随机变量. 令 $X_n = Z + \zeta_n$. 由均方大数定理, 我们有

$$\begin{aligned}Z &= Z + \lim_{m\to\infty} \frac{1}{m}(\zeta_n + \cdots + \zeta_{n-m+1}) \\ &= \lim_{m\to\infty} \frac{1}{m}(X_n + \cdots + X_{n-m+1}),\end{aligned}$$

所以在已知所有过去的情况下 Z 是一个允许预测量. 显然它还是最佳的预测量. 但 Z 不能够表示为 (4.3) 的形式. 然而, 我们知道最小均方误差预测量 $\widehat{X}^*$ 是存在的. 在大部分实际关心的情况中, 这样一个预测量有表达式

$$\widehat{X}_n^* = \sum_{k=1}^{\infty} \alpha_k^* X_{n-k}, \tag{4.4}$$

因此, 花些时间详细研究这种情况是值得的. 由于对 $X_{n-1}, X_{n-2}, \cdots$ 中任意有限线性组合 $U, E[(X_n - \widehat{X}_n^*)U] = 0$, 我们必有

$$E[(X_n - \widehat{X}_n^*)X_{n-k}] = 0, \quad k = 1, 2, \cdots. \tag{4.5}$$

由此推得对每个 U,

$$E[(X_n - \widehat{X}_n^*)U] = 0, \tag{4.6}$$

这里 $U = u_1X_{n-1} + u_2X_{n-2} + \cdots + u_pX_{n-p}$, 并由叉积的连续性, (4.6) 对于所有的

472

允许预测量 U 成立. 这样 (4.5) 提供了预测量 (4.4) 是最佳的一个充分必要条件.

现在我们把这个准则应用于 9.2 节的自回归过程. 假设 $\{X_n\}$ 是满足

$$X_n = \lambda_1X_{n-1} + \cdots + \lambda_pX_{n-p} + \xi_n \tag{4.7}$$

的协方差平稳过程, 其中 $\lambda_1, \cdots, \lambda_p$ 是固定的, $\{\xi_n\}$ 是一列具有共同方差和零均值的不相关随机变量. 把 (4.7) 写成

$$X_n = \widehat{X}_n + \xi_n,$$

其中,

$$\widehat{X}_n = \lambda_1X_{n-1} + \cdots + \lambda_pX_{n-p}.$$

在附注 2.1 中我们指出, 若

$$x^p - \lambda_1x^{p-1} - \cdots - \lambda_p = 0$$

的所有根 $x_1, x_2, \cdots, x_p$ 有 $|x_i| < 1$, 则 ξ_n 与 $X_{n-k}(k \geqslant 1)$ 不相关. 由此推得 $\widehat{X}_n$ 是 X_n 在已知所有过去条件下的最小均方线性预测量, 因为

$$E[(X_n - \widehat{X}_n)X_{n-k}] = E[\xi_nX_{n-k}] = 0, \quad k = 1, 2, \cdots,$$

即满足条件 (4.5).

回到一般情况, 假设存在一个形如

$$\widehat{X}_n^* = \sum_{k=1}^{\infty} \alpha_k^* X_{n-k} \tag{4.8}$$

的最佳预测量. 如前面指出的一样, 这不一定成立. 但是, 由于叉积在均方极限下是连续的, 将其代入 (4.5), 并交换期望与和号, 推得

$$R(k) = \sum_{l=1}^{\infty} \alpha_l^* R(k-l), \quad k = 1, 2, \cdots. \tag{4.9}$$

如果这个方程有解 $(\alpha_1^*, \alpha_2^*, \cdots)$, 且使得所对应的级数 (4.8) 依均方收敛, 则 (4.8) 确定一个最小均方误差预测量.

例 假设 $\{X_n, n = 0, \pm 1, \cdots\}$ 是具有协方差函数

$$R(u) = \begin{cases} 1, & u = 0, \\ \lambda/(1+\lambda^2), & |u| = 1, \\ 0, & |u| > 1 \end{cases}$$

473

的协方差平稳过程, 其中 $0 < \lambda < 1$. 方程 (4.9) 变成

$$\lambda = (1+\lambda^2)\alpha_1^* + \lambda\alpha_2^*,$$
$$0 = \lambda\alpha_{k-1}^* + (1+\lambda^2)\alpha_k^* + \lambda\alpha_{k+1}^*, \quad k = 2, 3, \cdots.$$

容易验证它的一个解是

$$\alpha_k^* = -(-\lambda)^k, \quad k = 1, 2, \cdots,$$

因此最小均方误差线性预测量是

$$\widehat{X}_n^* = \lambda X_{n-1} - \lambda^2 X_{n-2} + \lambda^3 X_{n-3} - \cdots.$$

有趣的是, 虽然 X_n 和 $X_{n-k} (k \geqslant 2)$ 是不相关的, 但 X_n 的最佳线性预测量与过程所有过去有关.

9.5 遍历理论和平稳过程

遍历定理指出随机过程在一定条件之下关于时间的平均

$$\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$$

当观测周期次数 n 很大时将收敛. 最重要的遍历定理是关于独立同分布且具有有限均值 $m = E[X_k]$ 的随机变量序列 $\{X_n\}$ 的强大数定理, 它断言样本平均 $\overline{X}_n$ 以概率 1 收敛于均值 m. 用符号表示即为

$$\Pr\{\lim_{n\to\infty} \overline{X}_n = m\} = 1.$$

平稳过程对推广大数定理提供了一个很自然的条件, 因为这个过程的均值是常数 $m = E[X_n]$, 不依赖于时间.

我们暂且考虑对任意一个过程估计其未知的均值函数 $m(n) = E[X_n]$ 的问题. 一般来说, 我们必须取过程的很大数目 N 个实现, 譬如说是

474

$$\begin{gathered}\{X_n^1; n = 1, 2, \cdots\},\\ \{X_n^2; n = 1, 2, \cdots\},\\ \vdots\\ \{X_n^N; n = 1, 2, \cdots\},\end{gathered}$$

然后计算它们的算术平均值

$$\overline{X}_{(n)} = \frac{1}{N}\{X_n^1 + \cdots + X_n^N\},$$

用它们作为 $m(n)$ 的估计量. 当然, 如果 $m(n)$ 是常数, 例如平稳过程即为此情况, 我们可以另外关于 n 个时间求平均, 并计算总平均值

$$\overline{\overline{X}} = \frac{1}{n}\{\overline{X}(1) + \cdots + \overline{X}(n)\}.$$

然而, 一般说来关键仍在于估计过程的均值需要整个过程不同的独立的实现. 作为比较, 让我们考虑对于服从遍历定理的平稳过程 $\{X_n\}$ 相同的估计问题. 据遍历定理,

$$\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n) \to m, \quad 当 \quad n \to \infty. \tag{5.1}$$

由 (5.1) 可知, 我们仅需要观测过程的单个实现, 但必须有足够长的时间长度. 然后利用 $\overline{X}_n$ 作为常数均值 m 的估计. 平稳过程的遍历理论的实际价值很大程度上正是由于这个事实. 对于这样的过程, 其均值 (和协方差函数) 通常可以通过过程的一个实现来估计.

本节我们将给出关于平稳过程的两个遍历定理. 如同强大数定理和弱大数定理, 遍历定理也有不同形式. 它们的区别在于假设和收敛方式. 其中第一个定理推广了弱大数定理, 适用于协方差平稳过程, 其收敛方式是按均方意义. 第二个定理推广了强大数定理, 它要求过程是严格平稳的, 其收敛方式是以概率 1.

均方遍历定理

设 $\{X_n\}$ 是协方差平稳过程, 其均值为 m, 协方差函数为

$$R(u) = E[\{X_{n+u} - m\}\{X_n - m\}].$$

令

$$\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n).$$

一种弱大数定理指出, 当 $\{X_n\}$ 不相关时, $\overline{X}_n$ 在均方意义下收敛于 m. 均方遍历定理得到相同的结果, 只假设 X_n 是渐近不相关的, 即协方差 $R(u)$ 当 u 增大时其 Cesaro 极限为 0.

475

定理 5.1 假设 $\{X_n\}$ 是具有协方差函数 $R(u)$ 的协方差平稳过程, 则

$$\lim_{N\to\infty}\frac{1}{N}\sum_{u=0}^{N-1}R(u)=0, \tag{5.2}$$

当且仅当

$$\lim_{N\to\infty}E[(\overline{X}_N-m)^2]=0. \tag{5.3}$$

证明 由于 $m=E[X_n]$, 我们可以把 (5.3) 看作 $\overline{X}_N$ 方差的极限, 而 (5.2) 看作 $\overline{X}_N$ 和 X_1 协方差的极限. 这样, 定理阐明了 $\overline{X}_N$ 的方差收敛于 0, 当且仅当 X_1 和 $\overline{X}_N$ 的协方差收敛于 0.

令

$$Y_n=X_n-m,\quad 且\quad \overline{Y}_N=\frac{1}{N}(Y_1+\cdots+Y_N),$$

然后, 利用施瓦兹不等式,

$$\begin{aligned}\left\{\frac{1}{N}\sum_{u=0}^{N-1}R(u)\right\}^2&=\{E[Y_1\overline{Y}_N]\}^2\\&\leqslant E[Y_1^2]\cdot E[\overline{Y}_N^2]\\&=R(0)\cdot E[(\overline{X}_N-m)^2].\end{aligned}$$

这样, (5.2) 由 (5.3) 推得. 为了证明相反的蕴涵关系, 我们设 (5.2) 成立. 现计算

$$\overline{Y}_N=\frac{1}{N}(Y_1+\cdots+Y_N)$$

的方差, 得

$$\begin{aligned}E[\overline{Y}_N^2]&=\frac{1}{N^2}E\left[\left(\sum_{k=1}^{N}Y_k\right)^2\right]\\&=\frac{1}{N^2}\left\{E\left[\sum_{k=1}^{N}Y_k^2+2\sum_{k<l}Y_kY_l\right]\right\}\\&=\frac{1}{N^2}\left\{NR(0)+2\sum_{l=1}^{N}\sum_{k=1}^{l-1}R(l-k)\right\}\\&=\frac{1}{N^2}\left\{2\sum_{l=1}^{N}\sum_{u=0}^{l-1}R(u)-NR(0)\right\}.\end{aligned}$$

476

注意到

$$\frac{1}{N}R(0) \to 0, \quad 当 \quad N \to \infty,$$

所以我们需要把注意力集中在第一项. 对于任意 $M < N$, 我们有

$$\frac{2}{N^2}\sum_{l=1}^{N}\sum_{u=0}^{l-1}R(u) = \frac{2}{N^2}\left\{\sum_{l=1}^{M}\sum_{u=0}^{l-1}R(u) + \sum_{l=M+1}^{N}\sum_{u=0}^{l-1}R(u)\right\}.$$

任给 $\varepsilon > 0$, 利用假设 (5.2), 选择 M 使得

$$\left|\frac{1}{l}\sum_{u=0}^{l-1}R(u)\right| \leqslant \varepsilon, \quad 如果 \quad l \geqslant M.$$

则

$$\left|\frac{2}{N^2}\sum_{l=M+1}^{N} l \times \frac{1}{l}\sum_{u=0}^{l-1}R(u)\right| \leqslant \frac{2}{N^2}\sum_{l=M+1}^{N} l\varepsilon \leqslant 2\varepsilon.$$

这样,

$$\left|\frac{2}{N^2}\sum_{l=1}^{N}\sum_{u=0}^{l-1}R(u)\right| \leqslant \frac{2}{N^2}\left|\sum_{l=1}^{M}\sum_{u=0}^{l-1}R(u)\right| + 2\varepsilon.$$

令 $N \to \infty$, 由于 M 是固定的, 右边第一项为 0. 又因为 ε 是任意正数, 我们必有

$$\lim_{N\to\infty}\frac{2}{N^2}\sum_{l=1}^{N}\sum_{u=0}^{l-1}R(u) = 0,$$

所以 $E[\overline{Y}_N^2] \to 0$. 这就完成了证明. ■

利用切比雪夫不等式

$$\Pr\{|\overline{X}_N - m| > \varepsilon\} \leqslant \frac{E[(\overline{X}_N - m)^2]}{\varepsilon^2}, \varepsilon > 0,$$

可得 (5.3) 蕴涵着对所有 $\varepsilon > 0$,

$$\lim_{N\to\infty}\Pr\{|\overline{X}_N - m| > \varepsilon\} = 0, \tag{5.4}$$

这样, 对足够大的 N, (5.4) 表明 $\overline{X}_N$ 依概率逼近于 m. 于是, 对未知的均值给出了一个估计.

477

如果相关函数对于充分大的 u 变得足够小, 即

$$\lim_{u\to\infty}R(u) = 0, \tag{5.5}$$

则 (5.2) 成立, 因为对任何 $M < N$, 我们有

$$\frac{1}{N}\sum_{u=0}^{N-1}R(u)=\frac{1}{N}\sum_{u=0}^{M-1}R(u)+\frac{1}{N}\sum_{u=M}^{N-1}R(u),$$

由于 (5.5), 选取 M 充分大右端第二项的绝对值可小于任意预先指定的 ε, 对于固定的 M, 右边第一项当 N 趋于无穷大时变为 0. 于是, (5.5) 是保证 $\overline{X}_N$ 收敛于均值 m 的一个充分条件.

假设 $\{X_n\}$ 是具有均值为 0 的协方差平稳过程. 相同的定理可以用来得到样本平均协方差

$$\widehat{R}_N(u)=\frac{1}{N}\sum_{k=0}^{N-1}X_kX_{k+u} \tag{5.6}$$

收敛于协方差

$$R(u)=E[X_nX_{n+u}]. \tag{5.7}$$

的条件. 如果我们记 $W_k=X_kX_{k+u}$, 则 (5.6) 变为

$$\widehat{R}_N(u)=\frac{1}{N}\sum_{k=0}^{N-1}W_k=\overline{W}_N.$$

这样, 如果 $\{W_k\}$ 构成一个协方差平稳过程, 并且

$$\lim_{T\to\infty}\frac{1}{T}\sum_{l=0}^{T-1}E\left[\{W_n-R(u)\}\cdot\{W_{n+l}-R(u)\}\right]=0, \tag{5.8}$$

则

$$\lim_{N\to\infty}E[(\widehat{R}_N(u)-R(u))^2]=0.$$

自然, 关于 $\{W_k\}$ 协方差所施加条件也就是对原来过程 $\{X_n\}$ 四阶乘积矩施加条件, 这就限制了上述结果的一般应用. 然而高斯过程是由它的均值和协方差函数确定的, 这样, 对于高斯过程 (5.8) 可以用协方差来表述. 条件 (5.8) 要求 478

$$\lim_{T\to\infty}\frac{1}{T}\sum_{l=0}^{T-1}E[\{X_nX_{n+u}-R(u)\}\cdot\{X_{n+l}X_{n+l+u}-R(u)\}]=0,$$

此式可归结为

$$\lim_{T\to\infty}\frac{1}{T}\sum_{l=0}^{T-1}\{E[X_nX_{n+u}X_{n+l}X_{n+l+u}]-R(u)^2\}=0.$$

而具有零均值的联合正态分布随机变量四阶乘积矩可由下式给出

$$E[X_iX_jX_kX_l] = \sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk},$$

其中 σ_{ij} 是 X_i 和 X_j 的协方差. 这时我们有

$$E[X_nX_{n+u}X_{n+l}X_{n+l+u}] = R(u)^2 + R(l)^2 + R(l-u)R(l+u).$$

这样, 在高斯过程的情况下, (5.8) 要求

$$\lim_{T\to\infty}\frac{1}{T}\sum_{l=0}^{T-1}\{R(l)^2 + R(l-u)R(l+u)\} = 0. \tag{5.9}$$

由于对任意两个实数 $a, b, |ab| \leqslant a^2 + b^2$, 所以

$$|R(l-u)R(l+u)| \leqslant R(l-u)^2 + R(l+u)^2.$$

于是

$$\lim_{T\to\infty}\frac{1}{T}\sum_{l=0}^{T-1}R(l)^2 = 0 \tag{5.10}$$

蕴涵着

$$\lim_{T\to\infty}\frac{1}{T}\sum_{l=0}^{T-1}R(l\pm u)^2 = 0,$$

因为这两个和至多只有有限项不同. 因此, 若 (5.10) 成立可推得 (5.9) 成立. 于是我们证明了下面的定理:

定理 5.2　设 $\{X_n\}$ 是具有协方差函数 $R(u)$ 且均值为 0 的高斯协方差平稳过程. 如果

479

$$\lim_{T\to\infty}\frac{1}{T}\sum_{u=0}^{T-1}R(u)^2 = 0,$$

则对任意固定的 $u = 0, \pm 1, \cdots,$

$$\lim_{T\to\infty}E[|\widehat{R}_T(u) - R(u)|^2] = 0,$$

其中 $\widehat{R}_T(u)$ 是样本协方差函数

$$\widehat{R}_T(u) = \frac{1}{T}\sum_{l=0}^{T-1}X_lX_{l+u}.$$

在介绍强收敛定理之前, 我们来证明一个一般的均方遍历定理.

定理 5.3 设 $\{X_n\}$ 是协方差平稳过程, 并定义一时间平均序列

$$\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$$

则存在一随机变量 $\overline{X}$, 它是 $\{\overline{X}_n\}$ 的均方极限,

$$\lim_{n\to\infty} \|\overline{X}_n - \overline{X}\| = 0.$$

证明 在证明之前, 我们先注意到, 一般来说, 极限 $\overline{X}$ 是随机的, 定理 5.1 给出一个使得 $\overline{X}$ 不是随机的而是一个常数 $m = E[X_1]$ 的充分必要条件.

依照均方收敛的柯西准则, 为证明定理只需证明

$$\lim_{m,n\to\infty} \|\overline{X}_n - \overline{X}_m\| = 0.$$

置

$$\mu_N = \inf \|\lambda_1 X_1 + \cdots + \lambda_N X_N\|,$$

其中下确界是对所有满足 $\lambda_1 + \cdots + \lambda_N = 1$ 的非负数 λ_i 取的. 注意到 $\mu_{N+1} \leqslant \mu_N$, 并令

$$\mu = \lim_{N\to\infty} \mu_N = \inf \mu_N.$$

不妨假设 $m < n$, 并计算

$$\begin{aligned}\|\overline{X}_m + \overline{X}_n\| &= \left\|\left(\frac{1}{m} + \frac{1}{n}\right) X_1 + \cdots + \left(\frac{1}{m} + \frac{1}{n}\right) X_m + \frac{1}{n} X_{m+1} + \cdots + \frac{1}{n} X_n\right\| \\ &= 2\|\lambda_1 X_1 + \cdots + \lambda_n X_n\| \geqslant 2\mu,\end{aligned}$$

480

其中

$$\lambda_i = \frac{1}{2}\left(\frac{1}{m} + \frac{1}{n}\right), \qquad i = 1, \cdots, m,$$

$$\lambda_i = \frac{1}{2n}, \qquad i = m+1, \cdots, n.$$

自然, 如果 $m > n$, 可得到相同不等式. 现在定理证明的关键是在于证明

$$\lim_{n\to\infty} \|\overline{X}_n\| = \mu, \tag{5.11}$$

因为, 如果我们能够做到这点, 由平行四边形法则和刚刚证明的不等式, 有

$$\begin{aligned}\|\overline{X}_n - \overline{X}_m\|^2 &= 2\|\overline{X}_n\|^2 + 2\|\overline{X}_m\|^2 - \|\overline{X}_n + \overline{X}_m\|^2 \\ &\leqslant 2\|\overline{X}_n\|^2 + 2\|\overline{X}_m\|^2 - (2\mu)^2 \\ &\leqslant 2\{|\|\overline{X}_n\|^2 - \mu^2| + |\|\overline{X}_m\|^2 - \mu^2|\},\end{aligned}$$

且 (5.11) 蕴涵着右边趋于 0.

这样, 我们可集中验证 (5.11). 任给正数 ε, 选择 N 和 $\lambda_1, \cdots, \lambda_N$ 使得

$$\|\lambda_1 X_1 + \cdots + \lambda_N X_N\| \leqslant \mu + \varepsilon,$$

自然, 其中 $\lambda_i \geqslant 0$ 且 $\lambda_1 + \cdots + \lambda_N = 1$. 令

$$Y_k = \lambda_1 X_k + \cdots + \lambda_N X_{k+N-1}.$$

则 $\{Y_k\}$ 是协方差平稳过程, 并且

$$\|Y_k\| \leqslant \mu + \varepsilon, \quad k = 1, 2, \cdots.$$

我们计算

$$\begin{aligned}
\overline{Y}_n &= \frac{1}{n}(Y_1 + \cdots + Y_n) \\
&= \frac{1}{n}(\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_N X_N \\
&\quad + \lambda_1 X_2 + \lambda_2 X_3 + \cdots + \lambda_N X_{N+1} + \cdots \\
&\quad + \lambda_1 X_n + \lambda_2 X_{n+1} + \cdots + \lambda_N X_{N+n-1}) \\
&= \lambda_1 \overline{X}_{1,n} + \lambda_2 \overline{X}_{2,n} + \cdots + \lambda_N \overline{X}_{N,n},
\end{aligned}$$

其中

481

$$\overline{X}_{k,n} = \frac{1}{n}(X_k + X_{k+1} + \cdots + X_{k+n-1}).$$

由于 $\overline{X}_n = \overline{X}_{1,n}, \lambda_1 - 1 = -(\lambda_2 + \cdots + \lambda_N)$, 我们有

$$\overline{Y}_n - \overline{X}_n = \lambda_2(\overline{X}_{2,n} - \overline{X}_{1,n}) + \cdots + \lambda_N(\overline{X}_{N,n} - \overline{X}_{1,n}).$$

然后利用三角不等式,

$$\|\overline{Y}_n - \overline{X}_n\| \leqslant \lambda_2 \|\overline{X}_{2,n} - \overline{X}_{1,n}\| + \cdots + \lambda_N \|\overline{X}_{N,n} - \overline{X}_{1,n}\|.$$

但是

$$\begin{aligned}
\|\overline{X}_{k,n} - \overline{X}_{1,n}\| &= \frac{1}{n}\|(X_k + \cdots + X_{k+n-1}) - (X_1 + \cdots + X_n)\| \\
&= \frac{1}{n}\|X_{n+1} + \cdots + X_{n+k-1} - X_1 - \cdots - X_{k-1}\| \\
&\leqslant \frac{1}{n}\{\|X_{n+1}\| + \cdots + \|X_{n+k-1}\| + \|X_1\| + \cdots + \|X_{k-1}\|\} \\
&= \frac{2k-2}{n}\|X_1\|, \qquad \text{对于} \quad k = 2, \cdots, N.
\end{aligned}$$

所以

$$\begin{aligned}\|\overline{Y}_n - \overline{X}_n\| &\leqslant \sum_{k=2}^{N} \lambda_k \frac{2k}{n} \|X_1\| \\ &\leqslant \frac{2N\|X_1\|}{n}.\end{aligned}$$

由此推得

$$\lim_{n\to\infty} \|\overline{Y}_n - \overline{X}_n\| = 0.$$

为完成 (5.11) 的证明, 注意到

$$\begin{aligned}\mu \leqslant \|\overline{X}_n\| &\leqslant \|\overline{X}_n - \overline{Y}_n\| + \|\overline{Y}_n\| \\ &\leqslant \|\overline{X}_n - \overline{Y}_n\| + \mu + \varepsilon,\end{aligned}$$

由于当 $n \to \infty$ 时 $\|\overline{X}_n - \overline{Y}_n\| \to 0$, 且 ε 是任意的, 我们必定有

$$\lim_{n\to\infty} \|\overline{X}_n\| = \mu.$$

这就验证了 (5.11), 从而完成了定理的证明. ■ 482

强遍历定理

定理 5.4 设 $\{X_n; n = 0, 1, 2, \cdots\}$ 是具有有限均值 $m = E[X_n]$ 的严格平稳过程. 记样本时间平均为

$$\overline{X}_n = \frac{1}{n}(X_0 + \cdots + X_{n-1}),$$

则序列 $\{\overline{X}_n\}$ 以概率 1 收敛于某个极限随机变量 $\overline{X}$, 即

$$\Pr\{\lim_{n\to\infty} \overline{X}_n = \overline{X}\} = 1.$$

证明 令

$$\overline{X}^* = \limsup_{n\to\infty} \overline{X}_n \quad 和 \quad \overline{X}_* = \liminf_{n\to\infty} \overline{X}_n.$$

序列 $\{\overline{X}_n\}$ 收敛事件自然就是事件 $\overline{X}^* = \overline{X}_*$, 并且它的余事件, 即 $\{\overline{X}_n\}$ 不收敛的事件是事件 $\overline{X}^* > \overline{X}_*$. 令 K 表示后面这个事件. 我们要证明

$$\Pr\{K\} = 0.$$

我们暂时先考虑使事件 K 出现的特别实现 $X_0 = x_0, X_1 = x_1, \cdots$, 则 $\overline{X}^* = \overline{x}^* > \overline{X}_* = \overline{x}_*$, 并且必定存在有理数 $\alpha < \beta$, 使得

$$\overline{x}^* > \beta > \alpha > \overline{x}_*.$$

由于仅存在可数多对有理数, 我们令 $(\alpha_1,\beta_1),(\alpha_2,\beta_2),\cdots$ 是所有这样性质的数对, 其中 $\alpha_k<\beta_k$. 令 K_k 是事件

$$\overline{X}^*>\beta_k>\alpha_k>\overline{X}_*.$$

如果事件 K 出现, 则对某个 k, 事件 K_k 必须出现. 上面我们刚刚证明了这一点. 反之, 如果某个 K_k 出现, 则事件 K 也出现. 这样, $K=\bigcup_{k=1}^{\infty}K_k$, 并且

$$\Pr\{K\}\leqslant\sum_{k=1}^{\infty}\Pr\{K_k\}.$$

由此推得, 为证明 $\Pr\{K\}=0$, 只需证明对于所有的 $k,\Pr\{K_k\}=0$. 即我们只需证明如果 α 和 β 是满足 $\alpha<\beta$ 的任意有理数,

483

$$\Pr\{\overline{X}^*>\beta>\alpha>\overline{X}_*\}=0.$$

令 A 表示事件 $\overline{X}^*>\beta>\alpha>\overline{X}_*,I(A)$ 表示事件 A 的示性随机变量:

$$I(A)=\begin{cases}1, & \text{如果 } A \text{ 出现},\\ 0, & \text{如果 } A \text{ 不出现}.\end{cases}$$

为证明 $\Pr\{A\}=E[I(A)]=0$, 我们首先要证明不等式

$$E[(X_0-\beta)I(A)]\geqslant 0 \tag{5.12}$$

和

$$E[(\alpha-X_0)I(A)]\geqslant 0. \tag{5.13}$$

把上两式相加得到

$$E[(\alpha-\beta)I(A)]\geqslant 0$$

或

$$(\alpha-\beta)\Pr\{A\}\geqslant 0.$$

由于 $\alpha<\beta$. 因此 $\Pr\{A\}=0$, 这就完成了证明. 下面我们的目标是建立不等式 (5.12) 和 (5.13).

显然 $I(A)=I(\limsup\overline{X}_n>\beta \quad 和 \quad \liminf\overline{X}_n<\alpha)$ 是整个序列 $X_0,X_1,\cdots$ 的函数. 回忆起实数列 $a_1,a\cdots$ 的上极限和下极限分别和任意位移序列 $a_{k+1},a_{k+2},\cdots$ 的上极限和下极限相等, 由此推得, 对于任意固定的 k,

$$I(A)=I(\limsup\overline{X}_{k+n}>\beta \quad 和 \quad \liminf\overline{X}_{k+n}<\alpha), \tag{5.14}$$

因此 $I(A)$ 在这个意义下关于时间移动是不变的这在后面证明中是很重要的.

引进记号 $Y_i = X_i - \beta, i = 0, 1, \cdots$ 且

$$S_{i,k} = Y_i + Y_{i+1} + \cdots + Y_{i+k-1}, \quad k \geqslant 1.$$

在 $S_{i,k}$ 中存在 k 个被加项, 初始项是 Y_i. 令

$$M_{i,n} = \max\{0, S_{i,1}, \cdots, S_{i,n}\}.$$

显然,

$$0 < S_{0,n} = Y_0 + \cdots + Y_{n-1} = (X_0 + \cdots + X_{n-1}) - n\beta,$$

当且仅当

$$\overline{X}_n = n^{-1}(X_0 + \cdots + X_{n-1}) > \beta.$$

484

若事件 A 出现, 则对某个正整数 n, 不等式 $\overline{X}_n > \beta$ 成立, 从而推得事件 $\{M_0, n > 0\}$ 对于足够大的 n 必须出现. 因此, 我们有

$$A \subset \bigcup_{n=1}^{\infty}\{M_{0,n} > 0\},$$

根据 $M_{0,n}$ 的单调性, 有

$$I(A) = \lim_{n\to\infty} I(A)I\{M_{0,n} > 0\},$$

其中, $I\{M_{0,n} > 0\}$ 照例表示事件 $\{M_{0,n} > 0\}$ 的示性随机变量. 最后借助于假设 $E[|Y_0|] < \infty$, 我们可以交换极限与期望, 因此得到

$$E[Y_0 I(A)] = \lim_{n\to\infty} E[Y_0 I\{M_{0,n} > 0\} I(A)].$$

我们现在来验证两个不等式 (5.12) 和 (5.13). 显然,

$$\begin{aligned} M_{1,n} &= \max\{0, S_{1,1}, \cdots, S_{1,n}\} \\ &\geqslant S_{1,k}, \qquad\qquad \text{对 } k = 1, \cdots, n-1, \end{aligned}$$

所以

$$Y_0 + M_{1,n} \geqslant Y_0 + S_{1,k} = S_{0,k+1}, \quad k = 1, \cdots, n-1.$$

改写为

$$Y_0 \geqslant S_{0,k} - M_{1,n}, \quad k = 2, \cdots, n.$$

由于 $M_{1,n} \geqslant 0$, 所以下式是平凡的

$$Y_0 \geqslant Y_0 - M_{1,n},$$

上面两个不等式结合得出

$$Y_0 \geqslant \max\{S_{0,1}, \cdots, S_{0,n}\} - M_{1,n},$$

由于 $M_{0,n} = \max\{0, S_{0,1}, \cdots, S_{0,n}\}$, 所以

$$Y_0 \geqslant M_{0,n} - M_{1,n}, \quad \text{如果} \quad M_{0,n} > 0.$$

这样,

$$\begin{aligned} & E[Y_0 I\{M_{0,n} > 0\} I(A)] \\ \geqslant & E[(M_{0,n} - M_{1,n}) I\{M_{0,n} > 0\} I(A)] \\ = & E[M_{0,n} I(A)] - E[M_{1,n} I\{M_{0,n} > 0\} I(A)] \\ \geqslant & E[M_{0,n} I(A)] - E[M_{1,n} I(A)], \end{aligned}$$

485

其中, 最后的不等式是由于 $M_{1,n} \geqslant 0$. 前面已指出事件 A 在时间移动下是不变的 [见 (5.14)] 由于过程是平稳的, $M_{0,n}$ 和 $M_{1,n}$ 具有相同分布, 因此 $E[M_{0,n} I(A)] = E[M_{1,n} I(A)]$. 于是, 在上面一串不等式的最左端式子为 0. 因此

$$E[Y_0 I(A)] = \lim_{n \to \infty} E[Y_0 I\{M_{0,n} > 0\} I(A)] \geqslant 0.$$

由于 $Y_0 = X_0 - \beta$, 我们已经证明了 (5.12), 即 $E[(X_0 - \beta) I(A)] \geqslant 0$. 对 $\widetilde{Y}_k = \alpha - X_k$ 应用上面证明方法得到 (5.13). 这就完成本定理的证明. ■

附注 5.1 在强遍历定理条件之下, 另外还成立

$$\lim_{n \to \infty} E[|\overline{X}_n - \overline{X}|] = 0. \tag{5.15}$$

因此

$$E[\overline{X}] = E[X_n] = m. \tag{5.16}$$

附注 5.2 进一步的初等考虑可以得到

$$\overline{X} = \lim_{n \to \infty} \frac{1}{n}(X_k + \cdots + X_{k+n-1}), \tag{5.17}$$

对每个 $k = 1, 2, \cdots$ 成立. 另外, 假设随机变量 $X_0, X_1, \cdots$ 是独立的. 由于 (5.17), 对每个 $k, \overline{X}$ 独立于 $X_0, \cdots, X_{k-1}$, 因此独立于 $\overline{X}_k$. 由于 $\overline{X} = \lim \overline{X}_k$, 我们必定有 $\overline{X}$ 独立于自己. 这样只有当 $\overline{X}$ 是常数时才有可能 (为什么?), 由 (5.16) 知道这个常数必是 m. 这样 $\lim\limits_{n \to \infty} \overline{X}_n = m$ 对于每个具有有限均值 m 的独立同分布随机变量序列均成立, 这就是强大数定理.

强大数定理促使我们寻找更加一般的条件使得极限随机变量 $\overline{X}$ 是一个常数. 在均方遍历定理中, 我们曾经得到, 当且仅当 $\overline{X}_n$ 和 $\overline{X}_1$ 是渐近不相关时 $\overline{X}$ 是一个常数. 在条件比较弱的强遍历定理中, 情况并不是这样简单, 然而某些原则可以确定.

如果 $x=(x_0,x_1,\cdots)$ 是实数列, 令 Tx 表示移位序列 $(x_1,x_2,\cdots)$. 我们称 T 是**移位算子**. 实数列构成的集合 A 称为**移位不变的**, 如果 Tx 是 A 的元素当且仅当 x 属于 A. 参看下面几个例子: 486

$$A_1=\{x: \text{对于某个 } k=1,2,\cdots,x_k=x_{k+1}=\cdots=0\},$$
$$A_2=\{x:\limsup x_n=a\},$$
$$A_3=\{x:\lim n^{-1}(x_1+\cdots+x_n)=b\}.$$

读者可以验证 A_1,A_2 和 A_3 都是移位不变的集合.

平稳过程称为**遍历的**, 若对任何一个移位不变集合 $A,\Pr\{(X_0,X_1,\cdots)\in A\}$ 或者是 0, 或者是 1.

定理 5.5 设 $\{X_n\}$ 是遍历平稳过程, 并具有有限均值 m, 则以概率 1

$$\lim_{n\to\infty}\frac{1}{n}(X_1+\cdots+X_n)=m.$$

证明 对每个实数 a, 定义

$$A=\{x=(x_0,x_1,\cdots);\lim_{n\to\infty}n^{-1}(x_1+\cdots+x_n)\leqslant a\}.$$

则 A 是移位不变的. 由此推得对每个常数 a

$$\begin{aligned}&\Pr\{(X_0,X_1,\cdots)\in A\}\\=&\Pr\{\lim n^{-1}(X_1+\cdots+X_n)\leqslant a\}\\=&\Pr\{\overline{X}\leqslant a\}=0 \quad \text{或} \quad 1,\end{aligned}$$

因此 $\overline{X}$ 一定是常数随机变量. 由 (5.16) 知, 此常数只能是 m. ■

当然, 我们希望有一些等价的且更易于理解的遍历性条件. 但在这里我们不能详述. 我们下面不加证明地列出一些结果.

定理 5.6 设 $\{X_n\}$ 是平稳过程, 下面条件是等价的

(a) $\{X_n\}$ 是遍历的;

(b) 对于每一个移位不变集合 A,

$$\Pr\{(X_0,X_1,\cdots)\in A\}=0 \quad \text{或} \quad 1;$$

(c) 对于由实数列 $(x_0, x_1, \cdots)$ 组成的每个集合 A

487

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}I\{(X_j, X_{j+1}, \cdots)\in A\} = \Pr\{(X_0, X_1, \cdots)\in A\};$$

(d) 对于每个 $k=1,2,\cdots$ 和由实数向量 $(x_0, x_1, \cdots, x_k)$ 组成的任一集合 A,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}I\{(X_j, \cdots, X_{j+k})\in A\} = \Pr\{(X_0, \cdots, X_k)\in A\};$$

(e) 对每个 k 和每个 $k+1$ 变量的函数,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\phi(X_j, \cdots, X_{j+k}) = E[\phi(X_0, \cdots, X_k)],$$

倘若上面期望存在;

(f) 对于实数列 $(x_0, x_1, \cdots)$ 的任一函数 ϕ,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\phi(X_j, X_{j+1}, \cdots) = E[\phi(X_0, X_1, \cdots)],$$

倘若上面期望存在.

在这些条件中, 极限的存在是强遍历定理的一个推论. 例如, 如果 ϕ 是 $k+1$ 个变量的函数, 且满足 $E[|\phi(X_0, \cdots, X_k)|]<\infty$, 则

$$Y_n = \phi(X_n, \cdots, X_{n+k}), \quad n=0,1,\cdots,$$

确定一个平稳过程, 对它应用强遍历定理, 推得 $\lim_{n\to\infty} n^{-1}\sum_{j=1}^{n}Y_j$ 存在. 由定理 5.6 的 (e) 知这个极限是一常数.

附注 5.3 当 $\{X_n\}$ 是遍历平稳的, 由定理 5.6 的等价条件, 显然可得, 对实序列的任何函数 ϕ, 序列

$$Y_n = \phi(X_n, X_{n+1}, \cdots), \quad n=1,2,\cdots$$

也构成一个遍历平稳过程.

一个平稳过程 $\{X_n\}$ 称为**混合的**(或**强混合的**), 如果对实数列的任意集合 A 和 B

$$\begin{aligned}&\lim_{n\to\infty}\Pr\{(X_1, X_2, \cdots)\in A \quad 且 \quad (X_{n+1}, X_{n+2}, \cdots)\in B\}\\ =&\Pr\{(X_1, X_2, \cdots)\in A\}\Pr\{(X_1, X_2, \cdots)\in B\}.\end{aligned}$$

488

混合是一种渐近独立的形式, 因此每个混合过程应该是遍历的. 我们下面来验证这

个命题. 设 $A=B$ 是一个移位不变集合, 则 $(X_n, X_{n+1}, \cdots)\in B$ 当且仅当 $(X_1, X_2\cdots)\in B$. 应用混合的性质得到

$$\begin{aligned}&\Pr\{(X_1,X_2,\cdots)\in A\}\\&=\Pr\{(X_1,X_2,\cdots)\in A,(X_n,X_{n+1},\cdots)\in A\}\\&\to[\Pr\{(X_1,X_2,\cdots)\in A\}]^2,\quad \text{当}\quad n\to\infty.\end{aligned}$$

这只有当 $\Pr\{(X_1,X_2,\cdots)\in A\}=0$ 或 1 才有可能, 所以过程是遍历的. 还存在着别的关于混合的等价条件. 例如, 对于任意 $k=1,2,\cdots$ 和 k 维空间中的集合 A,B,

$$\begin{aligned}&\lim_{n\to\infty}\Pr\{(X_1,\cdots,X_k)\in A\quad \text{且}\quad (X_{n+1,\cdots,X_{n+k}})\in B\}\\&=\Pr\{(X_1,\cdots,X_k)\in A\}\Pr\{(X_1,\cdots,X_k)\in B\}.\end{aligned}$$

最后我们必须指出以上我们所叙述的还不能看作就是遍历理论, 实际上仅是遍历理论对平稳随机过程的一些应用. 遍历理论的现代观点是相当一般的. 令 L 是由可定义某种积分概念 $\int f$ 的函数 f 构成的抽象空间, 而 T 是把 L 中函数 f 映射于 L 中函数 Tf 的算子. 遍历理论是研究算子 T 的叠算子 T^n 的有关极限性质, 其中算子 T 满足:

(i) 如果 f 是非负函数, 则 Tf 也是非负函数;

(ii) $\int|Tf|\leqslant\int|f|$.

在平稳过程 $\{X_n\}$ 的情况下, L 是由所有满足 $E[|\phi(X_0,X_1,\cdots)|]<\infty$ 的函数 $f(\{X_n\})=\phi(X_0,X_1,\cdots)$ 构成的, 积分的概念是期望, $\int f(\{X_n\})=E[\phi(X_0, X_1,\cdots)]$, 而 T 是如下定义的移位算子:

$$T\phi(X_0,X_1,X_2,\cdots)=\phi(X_1,X_2,\cdots).$$

9.6 遍历理论的应用

在上一节开始时我们就已指出遍历定理表明时间平均 $\overline{X}_n=n^{-1}(X_0+\cdots+X_{n-1})$ 可作为对应过程期望 $m=E[X_0]$ 的估计. 关于遍历定理还有大量其他的应用. 我们下面仅列举三个例子.

A. 随机环境的分支过程

分支过程我们已在第 2 章例 F 中介绍过, 并在第 8 章作了比较详细的讨论, 该过程最重要的特点是个体后代分布关于时间是不变的. 但在许多情况下, 后代的分

布依赖于环境的改变. 作为这一领域最近工作的一个介绍, 我们来讨论当环境依照平稳遍历过程随机变化时的绝灭问题.

我们假定一个平稳遍历过程 $\bar{\zeta} = \{\zeta_n\}_{n=0}^{\infty}$, 称为**环境过程**. 每个 ζ_n 值对应于一个后代分布 $P_{\zeta n} = \{p_n(i)\}_{i=0}^{\infty}$. 显然, $p_n(i) \geqslant 0$ 且 $\sum_{i=0}^{\infty} p_n(i) = 1$ 则 $\{p_{\zeta_n}\}_{n=0}^{\infty}$ 也是一个平稳遍历过程, 显然它的值是离散概率分布.

设 $Z_n(= Z(n))$ 表示在第 n 代存在的个体数, 并取 $Z_0 = 1$. 在所规定的环境 $\bar{\zeta} = \{\zeta_n\}$ 下, 我们假设 $\{Z_n\}$ 除了第 n 代后代分布是 $p_{\zeta n}$ 外, 其余都按普通分支过程演变. 准确地说, 我们假设

$$E[s^{Z(n+1)}|\zeta_0, \cdots, \zeta_{n;} Z_0, \cdots, Z_n] = [\phi_{\zeta_n}(s)]^{Z(n)},$$
$$n = 0, 1, 2, \cdots, |s| \leqslant 1,$$

其中

$$\phi_{\zeta_n}(s) = \sum_{j=0}^{\infty} p_n(j) s^j$$

是对应于 p_{ζ_n} 的概率母函数. 换句话说, 在过去的环境和群体总量条件下, $Z(n+1)$ 是 $Z(n)$ 窝 (或胎) 后代的总和, 其中每窝的数目是独立同分布随机变量, 其共同分布是 p_{ζ_n}.

设 B_n 是事件 $Z_n = 0$. 显然 $B = \bigcup_{n=1}^{\infty} B_n$ 表示最终绝灭事件. 引进记号

$$q(\bar{\zeta}) = \Pr\{B|\bar{\zeta}\}$$

和

$$q = \Pr\{B\} = E[q(\bar{\zeta})].$$

在环境过程 $\bar{\zeta}$ 的条件下, $\{Z_n\}$ 像一个普通的分支过程, 只不过是**非时间齐次的**, 利用第 8 章的技巧容易导出公式[1]

490

$$E[s^{Z(n+1)}|\bar{\zeta}] = [\phi_{\zeta_0}\phi_{\zeta_1}(\cdots(\phi_{\zeta_n}(s))\cdots)].$$

为简洁起见, 当不会发生混淆时, 我们经常把 $\phi_{\zeta i}$ 记为 ϕ_i. 显然

$$q(\bar{\zeta}) = \lim_{n\to\infty} \phi_0(\phi_1(\cdots(\phi_n(0))\cdots))$$
$$= \phi_{\zeta_0}[\lim_{n\to\infty} \phi_1\phi_2(\cdots(\phi_n(0)))\cdots)],$$

或

$$q(\bar{\zeta}) = \phi_{\zeta_0}[q(T\bar{\zeta})], \tag{6.1}$$

1. 我们始终假设对任何 Z, 与其相关联的 $\phi_{\zeta_i}(s)$ 却不是一个常数函数.

其中 $T\overline{\zeta}$ 表示移位了的环境序列 $\{\zeta_1, \zeta_2, \cdots\}$. 此重要的函数方程 $q(\overline{\zeta}) = \phi_{\zeta_0}[q(T\overline{\zeta})]$ 是随机的, 类似于第 8 章的方程 $s = \phi(s)$.

考虑事件 $\{q(\overline{\zeta}) = 1\}$. 也就是说, 当随机环境导致群体必然灭绝时此事件发生. 我们在基本方程 $q(\overline{\zeta}) = \phi_{\zeta_0}[q[T\overline{\zeta}]]$ 中利用概率母函数 ϕ_ζ 的性质推得只要 $q(\overline{\zeta}) = 1$ 就有 $q(T\overline{\zeta}) = 1$. 即, 事件 $\{q(\overline{\zeta}) = 1\}$ 是移位不变的. 依假设, 环境过程 $\overline{\zeta} = \{\zeta_n\}$ 是遍历的, 所以我们有

$$\Pr\{q(\overline{\zeta}) = 1\} = 0 \quad 或 \quad 1. \tag{6.2}$$

令 $m_{\zeta_n} = \sum_{j=0}^{\infty} j p_n(j)$ 是在给定环境下第 n 代后代分布的条件均值. $\{m_{\zeta_n}\}$ 仍然是一个平稳遍历过程, 并且若

$$E[|\ln m_{\zeta_0}|] < \infty, \tag{6.3}$$

强遍历定理告诉我们

$$\lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} \ln m_{\zeta_k} = E[\ln m_{\zeta_0}].$$

现在我们假定 (6.3), 并附加条件

$$\Pr\{q(\overline{\zeta}) < 1\} = 1, \tag{6.4}$$

看看究竟能得到什么结论. 由于概率母函数是递增的且是凸的, 所以

$$m_{\zeta_0} = \phi'_{\zeta_0}(1) \geqslant \frac{1 - \phi_{\zeta_0}(s)}{1 - s}, \quad 对于任意 \quad s \in [0, 1). \tag{6.5}$$

依照 (6.4), $q(\overline{\zeta}) < 1$, 再由基本函数方程 $q(\overline{\zeta}) = \phi_{\zeta_0}[q(T\overline{\zeta})]$, 与 (6.4) 等价我们可约定 $q(T\overline{\zeta}) < 1$(见前页脚注). 由 (6.5) 及 (6.1) 我们得到

491

$$m_{\zeta_0} \geqslant \frac{1 - \phi_{\zeta_0}[q(T\overline{\zeta})]}{1 - q(T\overline{\zeta})} = \frac{1 - q(\overline{\zeta})}{1 - q(T\overline{\zeta})}.$$

由平稳性质可以把上式推广为

$$m_{\zeta_n} \geqslant \frac{1 - q(T^n\overline{\zeta})}{1 - q(T^{n+1}\overline{\zeta})},$$

其中 $T^n\overline{\zeta}$ 是序列 $(\zeta_n, \zeta_{n+1}, \cdots)$. 由此有

$$\begin{aligned}\frac{1}{n}\sum_{k=0}^{n-1} \ln m_{\zeta_k} &\geqslant \frac{1}{n}\sum_{k=0}^{n-1} \ln\left\{\frac{1 - q(T^k\overline{\zeta})}{1 - q(T^{k+1}\overline{\zeta})}\right\} \\ &= \frac{1}{n}[\ln(1 - q(\overline{\zeta})) - \ln(1 - q(T^n\overline{\zeta}))] \\ &\geqslant \frac{1}{n}\ln[1 - q(\overline{\zeta})].\end{aligned}$$

如果 $E[|\ln m_{\zeta_0}|] < \infty$, 正如我们所假设的, 则由强遍历定理, 左边收敛于 $E[\ln m_{\zeta_0}]$, 而右边为 0 因此

$$E[\ln m_{\zeta_0}] \geqslant 0.$$

下面证明情况 $E[\ln m_{\zeta_0}] = 0$ 不可能出现. 为减少证明的难度, 我们假设

$$E\{|\ln[1-q(\overline{\zeta})] - \ln[1-q(T\overline{\zeta})]|\} < \infty.$$

于是

$$Y_n = \ln[1-q(T^n\overline{\zeta})] - \ln[1-q(T^{n+1}\overline{\zeta})]$$

构成一个遍历平稳过程, 对这个过程应用遍历定理, 有

$$\begin{aligned} E\left[\ln \frac{1-q(\overline{\zeta})}{1-q(T\overline{\zeta})}\right] &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \ln\left(\frac{1-q(T^k\overline{\zeta})}{1-q(T^{k+1}\overline{\zeta})}\right) \\ &= \lim_{n\to\infty} \frac{1}{n}\{\ln[1-q(\overline{\zeta})] - \ln[1-q(T^n\overline{\zeta})]\} \\ &\geqslant 0. \end{aligned} \tag{6.6}$$

由 (6.5) 我们知道

492

$$\ln \phi'_{\zeta_0}(1) - \ln\left\{\frac{1-\phi_{\zeta_0}[q(T\overline{\zeta})]}{1-q(T\overline{\zeta})}\right\} \geqslant 0.$$

如果我们假设 $0 = E[\ln m_{\zeta_0}] = E[\ln \phi'_{\zeta_0}(1)]$, 则不等式

$$0 \leqslant -E\left[\ln\left\{\frac{1-\phi_{\zeta_0}[q(T\overline{\zeta})]}{1-q(T\overline{\zeta})}\right\}\right]$$

成立. 与 (6.6) 比较, 我们推得

$$E\left[\ln\left\{\frac{1-\phi_{\zeta_0}[q(T\overline{\zeta})]}{1-q(T\overline{\zeta})}\right\}\right] = 0,$$

因此唯一值是 $q(T\overline{\zeta}) = 1$, 这与规定 $\Pr\{q(\overline{\zeta}) < 1\} = 1$ 相违背, 由此 $E[\ln m_{\zeta_0}] > 0$ 一定成立.

综上所述, 如果 $E[\ln m_{\zeta_0}] < \infty$, 则 $\Pr\{q(\overline{\zeta}) < 1\} = 1$ 蕴涵着 $E[\ln m_{\zeta_0}] > 0$. 相反, 由 $E[\ln m_{\zeta_0}] \leqslant 0$ 可得 $\Pr\{q(\overline{\zeta}) = 1\} = 1$.[利用 (6.2)]

B. 随机游动的幅度

设 $X_1, X_2, \cdots$ 是独立同分布随机变量, 其可能值是整数 $0, \pm1, \pm2, \cdots$. 考虑部分和过程 $S_n = X_1 + \cdots + X_n, n \geqslant 1$, 及 $S_0 = 0$. 定义在前 n 个和中的幅度 R_n 为 $\{S_1, \cdots, S_n\}$ 中不同数值的个数. 我们有如下表达式

$$R_n = \sum_{k=1}^{n} I_k,$$

其中

$$I_k = \begin{cases} 1, & \text{如果} \quad S_j \neq S_k, \quad \text{对于} j = 1, \cdots, k-1, \\ 0, & \text{其余}. \end{cases}$$

现在

$$\begin{aligned} E[I_k] &= \Pr\{S_k - S_{k-1} \neq 0, S_k - S_{k-2} \neq 0, \cdots, S_k - S_1 \neq 0\} \\ &= \Pr\{X_k \neq 0, X_k + X_{k-1} \neq 0, \cdots, X_k + \cdots + X_2 \neq 0\} \\ &= \Pr\{S_1 \neq 0, S_2 \neq 0, \cdots, S_{k-1} \neq 0\} \end{aligned}$$

(因为 X_i 是独立同分布的). 我们得到

$$\lim_{k\to\infty} E[I_k] = \Pr\{S_n \neq 0 \quad \text{对于所有的} \quad n \geqslant 1\}. \tag{6.7}$$

注意到 $\{S_n \neq 0$ 对于所有的 $n \geqslant 1\}$ 是过程 $\{S_n\}$ 未回到原点 $S_0 = 0$ 的事件. 但每个收敛序列又在 Cesaro 平均意义下收敛. 由 (6.7) 有

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} E[R_n] &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} E[I_k] \\ &= \Pr\{S_n \neq 0 \quad \text{对于所有的} \quad n \geqslant 1\}. \end{aligned} \tag{6.8}$$

493

若把上面左边的期望去掉, 极限式子仍然成立, 即以概率 1 成立

$$\lim_{n\to\infty} \frac{R_n}{n} = \Pr\{S_n \neq 0 \quad \text{对于所有的} \quad n \geqslant 1\} \tag{6.9}$$

我们下面通过验证与其等价的两个关于上、下极限的不等式来证明这个事实. 首先用 m 表示任意正整数, 并令 Z_k 是从 $(k-1)m+1$ 至 km 期间和序列 $\{S_i\}$ 经过的不同点的数目, 等价地说, Z_k 是在集合 $\{S_{(k-1)m+1}, \cdots, S_{km}\}$ 中的幅度. 注意到 Z_k 仅依赖于下标 n 介于 $(k-1)m+1$ 和 km 之间的 X_n, 所以 Z_k 是彼此独立的随机变量, $|Z_k| \leqslant m$, 显然 Z_k 也是同分布的. 验证不等式 $R_{km} \leqslant Z_1 + \cdots + Z_k$ 并应用大数定理得到

$$\limsup_{k\to\infty} \frac{R_{km}}{km} \leqslant \limsup_{k\to\infty} \frac{Z_1 + \cdots + Z_k}{mk} = \frac{E[Z_1]}{m}.$$

用 $[n/m]$ 表示不超过 n/m 的最大整数, 我们有 $R_n \leqslant R_{([n/m]+1)m}$, 因此

$$\begin{aligned} &\limsup_{n\to\infty} \frac{1}{n} R_n \\ &\leqslant \limsup_{n\to\infty} \frac{([n/m]+1)m}{n} \frac{R_{([n/m]+1)m}}{([n/m]+1)m} \\ &= \frac{E[Z_1]}{m}. \end{aligned}$$

上式对任何 m 成立. 但 $Z_1 = R_m$, 故由 (6.8) 有

$$\limsup_{n\to\infty} \frac{1}{n} R_n \leqslant \limsup_{n\to\infty} \frac{E[R_n]}{n} = \Pr\{S_n \neq 0 \quad \text{对于所有的} \quad n \geqslant 1\}. \tag{6.10}$$

相反不等式的证明比较困难, 我们需要强遍历定理.

定义

$$V_k = \begin{cases} 1, & \text{如果} \quad S_j \neq S_k, \quad \text{对于所有的} \quad j > k, \\ 0, & \text{其他}. \end{cases}$$

即如果 S_k 是时间 k 之后永远不会再到达的状态则 V_k 为 1, 否则为 0, 并且 $V_1 + \cdots + V_n$ 是至时刻 n 为止所到达而以后永远不再到达的状态的数目. 由于 R_n 是在 1 至 n 时刻所到达的而在时刻 $n+1$ 之前没有重复到达过的状态数目. 显然 $R_n \geqslant V_1 + \cdots + V_n$. 由于 $\{X_k\}$ 是平稳遍历过程 (见问题 21) 并且由附注 5.3, $\{V_k\}$也是, 因为

494

$$V_k = \begin{cases} 1, & \text{如果} \quad X_{k+1} \neq 0, X_{k+1} + X_{k+2} \neq 0, \cdots \\ 0, & \text{其他} \end{cases}$$
$$= \phi(X_{k+1}, X_{k+2}, \cdots),$$

其中

$$\phi(x_1, x_2, \cdots) = \begin{cases} 1, & \text{如果} \quad x_1 \neq 0, x_1 + x_2 \neq 0, x_1 + x_2 + x_3 \neq 0, \cdots \\ 0, & \text{其他}. \end{cases}$$

由遍历定理推得

$$\liminf_{n\to\infty} \frac{R_n}{n} \geqslant \liminf_{n} \frac{V_1 + \cdots + V_n}{n} = E[V_1]. \tag{6.11}$$

当然,

$$E[V_1] = \Pr\{S_k \neq 0, \quad \text{对所有} \quad k \geqslant 1\}.$$

结合 (6.10) 和 (6.11), (6.9) 得到证明.

C. 熵

概率是对单一事件的发生不确定性的度量. 熵是对事件集合不确定性的度量. 设 X 是以概率 p_i 取值 $i(i = 1, 2, \cdots, n)$ 的随机变量. X 的熵 (或事件 $\{X = i\}, i = 1, \cdots, n$ 的熵) 则按下式计算

$$H(X) = -\sum_{i=1}^{n} p_i \ln p_i \tag{6.12}$$

(我们规定 $0\cdot\ln 0=0$). 读者可以验证此定义具有如下三个性质: (i) 常数随机变量的熵是 0; (ii) 如果我们对随机变量附加上以概率 $p_{i+1}=0$ 取值 $i+1$, 其熵是不变的; (iii) 当 $p_1=\cdots=p_n=1/n$ 时, 熵取最大值 $\ln n$. 最后这一性质与我们关于不确性的直观是一致的. 例如随机变量

$$X_1=\begin{cases}1, & 以概率\quad 0.999,\\ 0, & 以概率\quad 0.001,\end{cases}$$

比随机变量

$$X_2=\begin{cases}1, & 以概率\quad \dfrac{1}{2},\\ 0, & 以概率\quad \dfrac{1}{2},\end{cases}$$

要肯定得多, 或易于预测得多.

495

我们可以把上述定义推广到两个随机变量的情况, 其中 $\Pr\{X=i,Y=j\}=p_{ij}$, 这时熵的公式是

$$H(X,Y)=-\sum_{i,j}p_{ij}\ln p_{ij}.$$

在已知 Y 条件下定义 X 的条件熵为

$$H(X|Y)=-\sum_j\Pr\{Y=j\}\sum_i p(i|j)\ln p(i|j),$$

其中 $p(i|j)=\Pr\{X=i|Y=j\}$. 利用等式 $p(i|j)=p_{ij}/\Pr\{Y=j\}$ 得到

$$\begin{aligned}H(X|Y)&=-\sum_{i,j}p_{ij}\ln[p_{ij}/\Pr\{Y=j\}]\\&=H(X,Y)-H(Y).\end{aligned}\tag{6.13}$$

由此得到重要关系式

$$H(X,Y)\geqslant H(Y).\tag{6.14}$$

换句话说, 两个随机变量的不确定性超过其中任何一个随机变量的不确定性.

证明在已知 Y 的条件之下 X 的不确定性不超过无条件下 X 的不确定性有点困难. 令 $q_j=\Pr\{Y=j\}$. 注意到函数 $\Gamma(t)=-t\ln t$. 当 $t>0$ 时是凹的, 利用 Jensen 不等式 (见 6.2 节, 凸性的广义定义) 得到

$$\begin{aligned}H(X|Y)&=\sum_i\sum_j q_j\Gamma[p(i|j)]\\&\leqslant\sum_i\Gamma\left[\sum_j q_jp(i|j)\right]\\&=H(X).\end{aligned}\tag{6.15}$$

由此式和 (6.13) 得到

$$H(X,Y) \leqslant H(X) + H(Y). \tag{6.16}$$

下面我们继续推广熵的定义. 如果 $X_1,\cdots,X_m$ 是具有联合分布的随机变量族, 并且 $\Pr\{X_1=i_1,\cdots,X_m=i_m\}=p(i_1,\cdots,i_m)$, 置

$$H(X_1,\cdots,X_m) = -\sum_{i_1,\cdots,i_m} p(i_1,\cdots,i_m)\ln p(i_1,\cdots,i_m).$$

我们有类似于 (6.13) 和 (6.15) 的结果,

$$H(X_1,\cdots,X_m) = H(X_1)+H(X_2|X_1)+H(X_3|X_1,X_2)+\cdots+H(X_m|X_1,\cdots,X_{m-1}). \tag{6.17}$$

和

496

$$H(X_k|X_1,\cdots,X_{k-1}) \leqslant H(X_k|X_2,\cdots,X_{k-1}). \tag{6.18}$$

现在令 $\{X_n\}$ 是平稳过程, 其中每个 X_k 的可能值是 $1,\cdots,N$. 由 (6.18) 及平稳性推得

$$\begin{aligned} H(X_{k-1}|X_1,\cdots,X_{k-2}) &= H(X_k|X_2,\cdots,X_{k-1}) \\ &\geqslant H(X_k|X_1,\cdots,X_{k-1}). \end{aligned}$$

由于序列是单调递减和非负的, 我们可以定义过程 $\{X_n\}$ 的熵为

$$H(\{X_n\}) = \lim_{k\to\infty} H(X_k|X_1,\cdots,X_{k-1}). \tag{6.19}$$

另一方面, 依照 (6.17), $(1/l)H(X_1,\cdots,X_l)$ 是 $H(X_k|X_1,\cdots,X_{k-1})$ 的 Cesaro 平均, 因此我们有

$$H(\{X_n\}) = \lim_{l\to\infty} \frac{1}{l} H(X_1,\cdots,X_l).$$

记 $p(i_1,\cdots,i_m)=\Pr\{X_1=i_1,\cdots,X_m=i_m\}$, 若 $\{X_n\}$ 是遍历的, 我们可得到出乎意外的结果, 即以概率 1,

$$\lim_{n\to\infty}\left[-\frac{1}{n}\ln p(X_0,\cdots,X_{n-1})\right] = H(\{X_n\}). \tag{6.20}$$

(6.20) 的右边是一常数, 可按照式 (6.19) 计算得到. 其左边部分是依随机序列 $\{X_n\}$ 的极限. 这个结果告诉我们, 对比较大的 n, 观察序列 $X_1,X_2,\cdots$ 的概率 $p(X_1,\cdots,X_n)$ 接近于 $\exp(-nH(\{X_n\}))$ 且以此为界.

我们先来验证当 $\{X_n\}$ 是平稳遍历有限马尔可夫链时, (6.20) 成立, 然后再来叙述并证明一般的结果.

令 $\boldsymbol{P}=\|P(i,j)\|_{i,j=1}^{N}$ 为不可约有限状态马尔可夫链 $\{X_n\}$ 的转移矩阵. 我们假设 $\Pr\{X_0=i\}=\pi(i)$, 其中 $\{\pi(i)\}_1^N$ 是对应于 $\boldsymbol{P}$ 的平稳分布. 计算

$$\begin{aligned}&H(X_n|X_0,\cdots,X_{n-1})\\&=-\sum\pi(i_0)P(i_0,i_1)\cdots P(i_{n-2},i_{n-1})\sum_{i_n}P(i_{n-1},i_n)\ln P(i_{n-1},i_n)\end{aligned}$$

(利用过程的马尔可夫链的特性和 $\pi(i)$ 是平稳分布)

$$=-\sum_{i,j}\pi(i)P(i,j)\ln P(i,j),$$

所以

$$H(\{X_n\})=-\sum_{i,j}\pi(i)P(i,j)\ln P(i,j).$$

497

而另一方面,

$$\begin{aligned}&-\frac{1}{n}\ln P(X_0,\cdots,X_{n-1})\\&=-\frac{1}{n}\ln\{\pi(X_0)P(X_0,X_1)\cdots P(X_{n-2},X_{n-1})\}\\&=\frac{1}{n}\sum_{i=0}^{n-2}W_i-\frac{1}{n}\ln\pi(X_0),\end{aligned}$$

其中

$$W_i=-\ln P(X_i,X_{i+1}).$$

一个不可约有限状态马尔可夫链若其初始状态服从平稳分布则生成一个遍历平稳过程. 由于仅有有限个状态, 因此 W_i 是有界的. 应用遍历定理得到

$$\begin{aligned}&\lim_{n\to\infty}\left[-\frac{1}{n}\ln P(X_0,\cdots,X_{n-1})\right]\\&=\lim_{n\to\infty}\left[-\frac{1}{n}\sum_{k=0}^{n-2}W_k\right]\\&=E[W_0]=-\sum_{i,j}\pi(i)P(i,j)\ln P(i,j)\\&=H(\{X_n\}),\end{aligned}$$

此即所要证明的. 下面是一般情形.

定理 6.1 设 $\{X_n\}$ 是具有有限状态空间 $\{1,\cdots,N\}$ 的遍历平稳过程. 令 $p(i_1,\cdots,i_m)=\Pr\{X_1=i_1,\cdots,X_m=i_m\}$ 和

$$H(\{X_n\})=\lim_{n\to\infty}\left(-\frac{1}{n}\sum_{k=1}^{n}\sum_{i_1,\cdots,i_k}p(i_1,\cdots,i_k)\ln p(i_1,\cdots,i_k)\right).$$

则以概率 1,

$$\lim_{n\to\infty}\left[-\frac{1}{n}\ln p(X_1,\cdots,X_n)\right]=H(\{X_n\}).$$

证明 不妨假设 $n=\cdots,-1,0,+1,\cdots$. 事实上, 给定任何单边平稳过程 $\{\widetilde{X}_n;n=0,1,\cdots\}$ 可确定双边平稳过程 $\{X_n;n=0,\pm1,\pm2,\cdots\}$, 使其有限维分布族与 $\{\widetilde{X}_n\}$ 的有限维分布族一致. 即令

$$\Pr\{X_m=i_m,\cdots,X_{m+k}=i_{m+k}\}=\Pr\{\widetilde{X}_1=i_m,\cdots,\widetilde{X}_{k+1}=i_{m+k}\},$$

498 对任何 $m=\cdots,-1,0,+1,\cdots$ 和 $k=1,2,\cdots$.

置

$$\begin{aligned}f_k(i;i_1,\cdots,i_k)&=\Pr\{X_k=i|X_0=i_k,\cdots,X_{k-1}=i_1\}\\&=\Pr\{X_0=i|X_{-k}=i_k,\cdots,X_{-1}=i_1\}\end{aligned}$$

(由平稳性). 由反向鞅收敛定理知存在极限

$$\begin{aligned}f(i;i_1,i_2,\cdots)&=\lim_{k\to\infty}f_k(i;i_1,\cdots,i_k)\\&=\Pr\{X_0=i|X_{-1}=i_1,X_{-2}=i_2,\cdots\}\\&=\Pr\{X_n=i|X_{n-1}=i_1,X_{n-2}=i_2,\cdots\}.\end{aligned}$$

置

$$g_0(i)=-\ln\Pr\{X_0=i\},$$

$$\begin{aligned}g_k(i_0,i_1\cdots,i_k)&=-\ln f_k(i_0;i_1,\cdots,i_k)\\&=-\ln\Pr\{X_0=i_0|X_{-k}=i_k,\cdots,X_{-1}=i_1\}\\&=-\ln\frac{\Pr\{X_{-k}=i_k,\cdots,X_{-1}=i_1,X_0=i_0\}}{\Pr\{X_{-k}=i_k,\cdots,X_{-1}=i_1\}},\end{aligned}$$

和

$$g(i,i_1,i_2,\cdots)=-\ln f(i;i_1,i_2,\cdots).$$

直接计算得到等式

$$-\frac{1}{n}\ln p(X_0,\cdots,X_{n-1})=\frac{1}{n}\sum_{k=0}^{n-1}g_k(X_k,\cdots,X_0).$$

如果我们能够用 $g(X_k,X_{k-1},\cdots)$ 代替 $g_k(X_k,\cdots,X_0)$ 则上式右边部分是典型的平均, 可应用遍历定理, 因为当 $\{X_k\}$ 是平稳过程时, $W_k=g(X_k,X_{k-1},\cdots)$ 也是平稳过程. 在马尔可夫过程情况下, 此代替是显然的, 因为由马尔可夫性, 对任意 $k\geqslant 1$ 都有 $g_k(X_k,\cdots,X_0)=g(X_k,X_{k-1},\cdots)$.

在当前情况下, 我们首先需要证明

$$E\left[\sup_{n\geqslant 1} g_n(X_n,\cdots,X_0)\right]<\infty. \tag{6.21}$$

固定 $\lambda>0$ 并分解如下事件

$$\begin{aligned}&\left\{\sup_{n\geqslant 1} g_n(X_n,\cdots,X_0)>\lambda\right\}\\&=\bigcup_{n=1}^{\infty}\{g_k(X_k,\cdots,X_0)\leqslant\lambda \quad 对于 \quad k=1,\cdots,n-1,\quad 且 \quad g_n(X_n,\cdots,X_0)>\lambda\}\\&=\bigcup_{n=1}^{\infty}\{(X_0,\cdots,X_n)\in E_n\},\end{aligned}$$

499

其中

$$E_n=\{(i_0,\cdots,i_n);\ g_k(i_k,\cdots,i_0)\leqslant\lambda \quad 对于 \quad k=1,\cdots,n-1 \quad 且 \quad g_n(i_n,\cdots,i_0)>\lambda\}.$$

我们已根据 $g_n(X_n,\cdots,X_0)>\lambda$ 成立的首次时间 n 分解了上面左边的事件. 由于这些事件是不相交的, 所以有

$$\begin{aligned}&\Pr\left\{\sup_{n\geqslant 1} g_n(X_n,\cdots,X_0)>\lambda\right\}\\&=\sum_{n=1}^{\infty}\Pr\{(X_0,\cdots,X_n)\in E_n\}.\end{aligned}$$

下面我们通过规定最后状态进一步分解事件. 令

$$E_n^{(i)}=\{(i_0,\cdots,i_{n-1});(i_0,\cdots,i_{n-1},i)\in E_n\}.$$

现在考虑

$$\begin{aligned}&\Pr\{(X_0,\cdots,X_n)\in E_n\}\\&=\sum_i\Pr\{(X_0,\cdots,X_{n-1})\in E_n^{(i)} \quad 且 \quad X_n=i\}.\end{aligned}\tag{6.22}$$

由条件概率的定义, 我们有

$$\begin{aligned}&\Pr\{X_n=i \quad 且 \quad (X_0,\cdots,X_{n-1})\in E_n^{(i)}\}\\&=\sum_{(i_0,\cdots,i_{n-1})\in E_n^{(i)}}\Pr\{X_n=i|X_0=i_0,\cdots,X_{n-1}=i_{n-1}\}\\&\quad\cdot\Pr\{X_0=i_0,\cdots,X_{n-1}=i_{n-1}\}.\end{aligned}\tag{6.23}$$

由于过程是平稳的, 我们有

$$\begin{aligned}&\Pr\{X_n=i|X_0=i_0,\cdots,X_{n-1}=i_{n-1}\}\\=&\Pr\{X_0=i|X_{-n}=i_0,\cdots,X_{-1}=i_{n-1}\}\\=&\exp[-g_n(i,i_{n-1},\cdots,i_0)].\end{aligned}$$

显然只要 $(i_0,\cdots,i_{n-1})$ 属于 $E_n^{(i)}, g_n(i,i_{n-1},\cdots,i_0)$ 一定超过 λ. 这样, (6.23) 中概率的界为

$$\begin{aligned}&\Pr\{X_n=i \text{ 和 } (X_0,\cdots,X_{n-1})\in E_n^{(i)}\}\\ \leqslant&\sum_{(i_0,\cdots,i_{n-1})\in E_n^{(i)}}\mathrm{e}^{-\lambda}\Pr\{X_0,=i_0,\cdots,X_{n-1}=i_{n-1}\}\\ \leqslant&\,\mathrm{e}^{-\lambda}\Pr\{(X_0,\cdots,X_{n-1})\in E_n^{(i)}\}.\end{aligned}$$

500

对每个固定 i, 集合 $E_1^{(i)},E_2^{(i)},\cdots$ 不相交, 其理由与 $E_1,E_2,\cdots$ 不相交是相同的, 把等式 (6.22) 关于 n 求和并利用所求的界, 有

$$\begin{aligned}&\sum_{n=1}^{\infty}\Pr\{(X_0,\cdots,X_n)\in E_n\}\\ =&\sum_i\sum_{n=1}^{\infty}\Pr\{(X_0,\cdots,X_{n-1})\in E_n^{(i)}\text{ 和 } X_n=i\}\\ \leqslant&\,\mathrm{e}^{-\lambda}\sum_i\left\{\sum_{n=1}^{\infty}\Pr\{(X_0,\cdots,X_{n-1})\in E_n^{(i)}\}\right\}\\ \leqslant&\,\mathrm{e}^{-\lambda}\sum_i 1\\ \leqslant&\,N\mathrm{e}^{-\lambda},\end{aligned}$$

其中 N 是过程不同状态 (i 值) 的数目. 这样我们证明了

$$\Pr\left\{\sup_{n\geqslant1}g_n(X_n,\cdots,X_0)>\lambda\right\}\leqslant N\mathrm{e}^{-\lambda}.$$

由此估计式易证得关系式 (6.21).

因为以概率 1,

$$g_k(X_0,X_{-1},\cdots,X_{-k})\to g(X_0,X_{-1},\cdots),\quad (k\to\infty)$$

并且左边函数族是一致可积的, 我们可以交换期望和极限号得到

$$E[g(X_0,X_{-1},\cdots)]=\lim_{k\to\infty}E[g_k(X_0,\cdots,X_{-k})]$$

$$= \lim_{n\to\infty} \frac{1}{n} E\left\{\sum_{k=0}^{n-1} g_k(X_k, \cdots, X_0)\right\}$$
$$= H(\{X_k\}).$$

第二个等号是由于平稳性.

注意,

$$\frac{1}{n}\sum_{k=0}^{n-1} g_k(X_k, \cdots, X_0)$$
$$= \frac{1}{n}\sum_{k=0}^{n-1} g(X_k, X_{k-1}, \cdots) + \frac{1}{n}\sum_{k=0}^{n-1}\{g_k(X_k, \cdots, X_0) - g(X_k, X_{k-1}, \cdots)\}. \quad (6.24)$$

我们已证明了 $E[g(X_k, X_{k-1}, \cdots)] < \infty$, 再由遍历定理推得

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} g(X_k, X_{k-1}, \cdots)$$
$$= E[g(X_0, X_{-1}, \cdots)] = H(\{X_n\}). \quad (6.25)$$

501

置

$$\phi_N(x_0, x_{-1}, \cdots) = \sup_{k\geqslant N} |g_k(x_0, \cdots, x_{-k}) - g(x_0, x_{-1}, \cdots)|,$$

和

$$Z_k^N = \phi_N(X_k, X_{k-1}, \cdots).$$

显然, $\{Z_k^N\}$ 是平稳遍历的且 $E[|Z_k^N|] < \infty$. 因此

$$\limsup_{n\to\infty} \left|\frac{1}{n}\sum_{k=0}^{n-1}\{g_k(X_k, \cdots, X_0) - g(X_k, X_{k-1}, \cdots)\}\right|$$
$$\leqslant \limsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1}\left|g_k(X_k, \cdots, X_0) - g(X_k, X_{k-1}, \cdots)\right|$$
$$\leqslant \limsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} Z_k^N = E[Z_1^N].$$

但当 $N\to\infty$ 时, $Z_1^N\to 0$, 并且可以验证极限号与期望可交换, 从而成立 $\lim_{N\to\infty} E[Z_1^N] = 0$. 由此推得式 (6.24) 右边第二项当 $n\to\infty$ 时趋于 0. 结合式 (6.25), 我们有

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} g_k(X_k, \cdots, X_0) = H(\{X_n\}),$$

这就是要证明的.

9.7 协方差平稳过程的谱分析

在本节将研究对于零均值协方差平稳过程使用调和函数的典型表示式. 前后我们曾猜想任意这样一个过程可表达为形如 9.1 节例 B 过程序列

$$X_n = \sum_{k=0}^{m}\{A_k \cos n\omega_k + B_k \sin n\omega_k\} \tag{7.1}$$

的平均平方极限, 其中 $0 \leqslant \omega_k \leqslant \pi$, 系数 $\{A_k\}$ 和 $\{B_k\}$ 互不相关且具有零均值和方差 $\sigma_k^2 = E[A_k^2] = E[B_k^2]$. 在 9.1 节中, 我们计算了 (7.1) 的协方差函数是

502

$$R(u) = \sigma^2 \sum_{k=0}^{m} p_k \cos u\omega_k, \tag{7.2}$$

其中, $\sigma^2 = \sigma_0^2 + \cdots + \sigma_m^2$, 且 $p_k = \sigma_k^2/\sigma^2$. 为了方便, 我们假设 $\omega_0 = 0$ 并把 (7.2) 写为对称形式

$$R(u) = \sigma^2 \sum_{k=-m}^{m} q_k \cos u\omega_k,$$

其中, $q_0 = p_0$, 且对于 $k = 1, \cdots, m, q_k = \dfrac{1}{2}p_k, q_{-k} = q_k, \omega_{-k} = -\omega_k$. 如 9.1 节中所述, 这提示我们作如下的推广:

$$R(u) = \sigma^2 \int_{-\pi}^{\pi} \cos u\omega \mathrm{d}F(\omega) \tag{7.3}$$

其中 F 是可能值在 $[-\pi, \pi]$ 上的对称随机变量的分布函数. 在本节我们将证明每个协方差函数可以写为 (7.3) 的形式, 其中 $F(\omega)$ 是某个唯一确定的概率分布函数. 我们把与协方差函数这样相联系的函数称为过程的**谱分布函数**.

一般来说, 从 (7.3) 出发可以进一步得到过程 $\{X_n\}$ 的典型表示式. 为粗略描述这个想法, 假设 $\{Z^{(1)}(\omega), 0 \leqslant \omega \leqslant \pi\}$ 和 $\{Z^{(2)}(\omega), 0 \leqslant \omega \leqslant \pi\}$ 是两个彼此不相关的随机过程, 并且各自在不相重迭区间上的增量也不相关. 进一步假设 $Z^{(i)}(0) = 0$ 及过程 $Z^{(i)}$ 具有共同的方差函数

$$\begin{aligned} E[Z^{(i)}(\omega)^2] &= \sigma^2[F(\omega) - F(-\omega)] \\ &= 2\sigma^2[F(\omega) - F(0)], \qquad 0 \leqslant \omega \leqslant \pi \end{aligned}$$

为说明方便, 在这个概略的介绍中, 我们还假设 $F(\omega)$ 是连续的.

令 $0 = \omega_0 < \omega_1 < \cdots < \omega_m < \omega_{m+1} = \pi$ 是给定的, 置 $\Delta Z_k^{(i)} = Z^{(i)}(\omega_{k+1}) - Z^{(i)}(\omega_k)$, 由于在不相重迭区间上增量是不相关的, 并且协方差函数是 F, 我们有

$$E[\Delta Z_k^{(i)} \Delta Z_l^{(j)}] = 0 \qquad 如果 \quad i \neq j \quad 或 \quad k \neq l,$$

且

$$\frac{1}{\sigma^2}E[(\Delta Z_k^{(i)})^2] = 2\{F(\omega_{k+1}) - F(\omega_k)\} = 2\Delta F_k.$$

让 $A_k = \Delta Z_k^{(1)}$ 和 $B_k = \Delta Z_k^{(2)}$, 等式 (7.1) 变为

$$\sum_{k=0}^{m} \cos n\omega_k \Delta Z_k^{(1)} + \sum_{k=0}^{m} \sin n\omega_k \Delta Z_k^{(2)},$$

503

它是黎曼 – 斯蒂尔切斯逼近和, 它在均方意义下收敛于如下积分:

$$\int_0^{\pi} \cos n\omega \mathrm{d}Z^{(1)}(\omega) + \int_0^{\pi} \sin n\omega \mathrm{d}Z^{(2)}(w).$$

关于高斯过程相应情况将在下一节研究.

协方差函数的谱表示

设 $\{X_n; n = 0, \pm 1, \pm 2, \cdots\}$ 是具有零均值和已知协方差函数的协方差平稳过程. 若需要, 可乘以一常数, 我们不妨假设过程的常数方差是 1. 这样, 我们所研究的协方差平稳过程 $\{X_n\}$, 其协方差函数

$$R(u) = E[X_n X_{n+u}], \quad u = 0, \pm 1, \pm 2, \cdots,$$

满足 $R(0) = 1$.

在叙述第一个定理之前, 我们给出两个简短定义. 首先, 一个实值函数 $R(u), u = 0, \pm 1, \pm 2, \cdots$ 称为**半正定的**, 如果对所有 $k = 1, 2, \cdots$ 和所有实数 $\alpha_1, \cdots, \alpha_k$,

$$\sum_{i=1}^{k}\sum_{j=1}^{k} = \alpha_i \alpha_j R(i-j) \geqslant 0. \tag{7.4}$$

其次, 一个随机变量 W, 或它的分布函数 F, 称为**对称的**, 如果 W 与 $-W$ 具有相同的分布. 这样, 如果 F 在所有连续点 ω 上满足 $F(\omega) = 1 - F(-\omega)$ 则 F 是对称的.

定理 7.1 设给定 $R(u), u = 0, \pm 1, \pm 2, \cdots$. 则下面命题等价:

(a) $R(u)$ 是具有零均值和方差为 1 的实值协方差平稳过程的协方差函数;

(b) $R(0) = 1$, 对于 $u = 0, 1, \cdots, R(u) = R(-u)$ 并且 $R(u)$ 是半正定的;

(c) 在 $[-\pi, \pi]$ 上存在一个对称概率分布函数 F, 使得

$$R(u) = \int_{-\pi}^{\pi} \cos u\omega \mathrm{d}F(\omega) \qquad u = 0, 1, \cdots. \tag{7.5}$$

504

证明 $(a) \Rightarrow (b)$. 假设 $R(u)$ 是具有零均值和方差为 1 的实值协方差平稳过程的协方差函数, 显然, $R(0) = E[X_n^2] = 1$, 并且 $R(u) = E[X_n X_{n+u}] = E[X_{n+u} X_n] =$

$R(-u)$. 为证明 $R(u)$ 是半正定的, 令 $\alpha_1, \cdots, a_k$ 是任意给定的实数, 计算

$$\begin{aligned} 0 \leqslant E\left[\left|\sum_{i=1}^{k} \alpha_i X_{n+i}\right|^2\right] &= E\left[\left(\sum_{i=1}^{k} \alpha_i X_{n+i}\right)\left(\sum_{j=1}^{k} \alpha_j X_{n+j}\right)\right] \\ &= E\left[\sum_{i=1}^{k}\sum_{j=1}^{k} \alpha_i\alpha_j X_{n+i}X_{n+j}\right] \\ &= \sum_{i=1}^{k}\sum_{j=1}^{k} \alpha_i\alpha_j E[X_{n+i}X_{n+j}] \\ &= \sum_{i=1}^{k}\sum_{j=1}^{k} \alpha_i\alpha_j R(i-j). \end{aligned}$$

这样由 (a) 推得 (b).

(b)⇒(a). 现在假设已知 $R(u)$ 具有 (b) 中所叙述的性质, 并置

$$a_{ij} = R(i-j).$$

则对任意 $k = 1, 2, \cdots$, 矩阵 $\boldsymbol{A} = \|a_{ij}\|_{i,j=1}^{k}$ 是对称的且半正定的. 根据第 1 章关于多元正态分布的说明, 任何这样一个矩阵对应于一个具有零均值且协方差阵为 A 的多元正态分布. 这个对应一致地确定任意有限集合 $(X_{n+1}, \cdots, X_{n+k})$ 的分布, 从而确定了过程 $\{X_n; n = 0, \pm 1, \cdots\}$ 的分布. 显而易见, 这时 R 就是该过程的协方差函数. 这样, 对于具有 (b) 中性质的 R, 至少存在一个协方差平稳过程, 以 R 为其协方差函数.

(c)⇒(b). 我们假设

$$R(u) = \int_{-\pi}^{\pi} \cos u\omega \mathrm{d}F(\omega) = E[\cos u\omega],$$

505 其中 W 是具有分布函数 F 的对称随机变量. 显然有 $R(0) = 1$ 和 $R(u) = R(-u)$.我们仅需要证明 $R(u)$ 是半正定的. 利用恒等式 $\cos(x-y) = \cos x \cos y + \sin x \sin y$, 有

$$\begin{aligned} \sum_{i=1}^{k}\sum_{j=1}^{k} \alpha_i\alpha_j R(i-j) &= \sum_{i=1}^{k}\sum_{j=1}^{k} \alpha_i\alpha_j E[\cos(i-j)W] \\ &= E\left[\sum_{i=1}^{k}\sum_{j=1}^{k} \alpha_i\alpha_j \cos(iW - jW)\right] \end{aligned}$$

$$= E\left[\left|\sum_{i=1}^{k}\alpha_i \cos iW\right|^2 + \left|\sum_{i=1}^{k}\alpha_i \sin iW\right|^2\right] \geqslant 0$$

这样便由 (c) 推得 (b).

(b)⇒(c). 设 R 是已知的. 对任意给定的 n 和 ω, 利用性质 (7.4), 令 $\alpha_j = \cos j\omega$, 得

$$0 \leqslant \sum_{i=1}^{n}\sum_{j=1}^{n} R(i-j)\cos i\omega \cos j\omega.$$

类似地, 令 $\alpha_i = \sin i\omega$, 得

$$0 \leqslant \sum_{i=1}^{n}\sum_{j=1}^{n} R(i-j)\sin i\omega \sin j\omega.$$

把这两个不等式相加并利用三角恒等式 $\cos(x-y) = \cos x\cos y + \sin x \sin y$ 推得

$$\begin{aligned}0 &\leqslant \frac{1}{2\pi n}\sum_{i=1}^{n}\sum_{j=1}^{n} R(i-j)\cos(i-j)\omega \\ &= \frac{1}{2\pi n}\sum_{u=-n+1}^{n-1}(n-|u|)R(u)\cos u\omega.\end{aligned}$$

如果我们定义

$$f_n(\omega) = \frac{1}{2\pi n}\sum_{u=-n+1}^{n-1}(n-|u|)R(u)\cos u\omega, \quad -\pi \leqslant \omega \leqslant \pi,$$

则

$$f_n(\omega) \geqslant 0.$$

其次计算

$$\begin{aligned}\int_{-\pi}^{\pi} f_n(\omega)\mathrm{d}\omega &= \frac{1}{2\pi n}\sum_{u=-n+1}^{n-1}(n-|u|)\left\{\int_{-\pi}^{\pi}\cos u\omega \mathrm{d}\omega\right\}R(u) \\ &= \frac{1}{2\pi n}(n)(2\pi)R(0) = 1.\end{aligned}$$

506

故, 对每个 n, f_n 是 $[-\pi, \pi]$ 上的概率密度函数. 由 f_n 的定义可知它是对称的:

$$f_n(\omega) = f_n(-\omega).$$

令 F_n 是对应于密度函数 f_n 的分布函数. 利用 f_n 的定义, 计算

$$\int_{-\pi}^{\pi}\cos u\omega \mathrm{d}F_n(\omega) = \int_{-\pi}^{\pi}\cos u\omega f_n(\omega)\mathrm{d}\omega$$

$$
\begin{aligned}
&= \int_{-\pi}^{\pi} \cos u\omega \left\{ \frac{1}{2\pi n} \sum_{k=-n+1}^{n-1} (n-|k|) \cos k\omega R(k) \right\} \mathrm{d}\omega \\
&= \sum_{k=-n+1}^{n-1} \left(1 - \frac{|k|}{n}\right) \frac{1}{2\pi} \left\{ \int_{-\pi}^{\pi} \cos u\omega \cos k\omega \mathrm{d}\omega \right\} R(k) \\
&= \sum_{k=-n+1}^{n-1} \left(1 - \frac{|k|}{n}\right) \delta_{uk} R(k) \\
&= \left(1 - \frac{|u|}{n}\right) R(u),
\end{aligned}
$$

其中当 $u = k$ 时 δ_{uk} 为 1, 否则为 0. 这样我们定义了一个分布函数序列 F_n, 对于它们, (7.5) 在下面意义下 "近似地" 是正确的,

$$
\left(1 - \frac{|u|}{n}\right) R(u) = \int_{-\pi}^{\pi} \cos u\omega \mathrm{d}F_n(\omega).
$$

由 Helly-Bray 引理 (第 1 章) 知, 存在分布函数 F 和子序列 $\{F_{n_k}\}$, 使得对每个有界连续函数 h,

$$
\lim_{k\to\infty} \int_{-\pi}^{\pi} h(w) \mathrm{d}F_{n_k}(\omega) = \int_{-\pi}^{\pi} h(\omega) \mathrm{d}F(\omega).
$$

与 $h(\omega) = \cos u\omega$ 时得到

$$
\begin{aligned}
\int_{-\pi}^{\pi} \cos u\omega \mathrm{d}F(\omega) &= \lim_{k\to\infty} \int_{-\pi}^{\pi} \cos u\omega \mathrm{d}F_{n_k}(\omega) \\
&= \lim_{k\to\infty} \left(1 - \frac{|u|}{n_k}\right) R(u) \\
&= R(u), \quad 对于 \quad u = 0, \pm 1, \cdots,
\end{aligned}
$$

507 这就完成了定理的证明. ■

按照等式 (7.5), 协方差函数 $R(u)$ 易由谱分布函数 F 确定. 事实上,

$$
R(u) = \int_{-\pi}^{\pi} \cos u\omega \mathrm{d}F(\omega)
$$

表明 $R(u)$ 是分布函数 $F(\omega)$ 的傅里叶 - 斯蒂尔切斯余弦变换. 反之, 一般地, 谱分布函数 F 可由对协方差函数 $R(u)$ 计算逆余弦变换得到. 我们下面不加证明地叙述当 $F(\omega)$ 可微分时的结果. $F(\omega)$ 的导数

$$
f(\omega) = \frac{\mathrm{d}F(\omega)}{\mathrm{d}\omega}, \qquad 对于 -\pi < \omega < \pi,
$$

称为**谱密度函数**. 在这种情况下,

$$R(u)=\int_{-\pi}^{\pi}\cos u\omega f(\omega)\mathrm{d}\omega$$

是谱密度函数 $f(\omega)$ 的余弦变换.

定理 7.2 设 $F(\omega)$ 是对应于协方差函数 $R(u)$ 的谱分布函数, 如果

$$\sum_{u=0}^{\infty}|R(u)|<\infty,$$

则 $F(\omega)$ 可微且其导函数为

$$f(\omega)=F'(\omega),\qquad -\pi<\omega<\pi.$$

这时有

$$\begin{aligned}f(\omega)&=\frac{1}{\pi}\left\{\frac{1}{2}R(0)+\sum_{u=1}^{\infty}\cos u\omega R(u)\right\}\\&=\frac{1}{2\pi}\sum_{u=-\infty}^{+\infty}R(u)\cos u\omega\\&=\frac{1}{2\pi}\sum_{u=-\infty}^{+\infty}\mathrm{e}^{\mathrm{i}u\omega}R(u).\end{aligned}$$

复值过程

为了理论和应用上需要, 我们下面来介绍复值平稳过程. 作为实例, 在通讯理论的许多领域中, 通常把一个交流电信号的瞬时电压和电流用复数来表示. 非常幸运, 这时我们可用很自然的方式达到预期的推广. [508]

令 $X=X_1+\mathrm{i}X_2$ 和 $Y=Y_1+\mathrm{i}Y_2$ 是复值随机变量, 我们定义其期望为

$$m_X=E[X]=E[X_1]+\mathrm{i}E[X_2],$$

协方差定义为

$$\operatorname{cov}[X,Y]=E[(X-m_X)\overline{(Y-m_Y)}],$$

其中 $\overline{(Y-m_Y)}$ 表示 $Y-m_Y$ 的共轭复数.

这样, 具有零均值复协方差平稳过程 $\{X_n\}$ 的协方差函数是复值函数

$$R(u)=E[X_n,\overline{X}_{n+u}],\qquad u=0,\pm1,\pm2,\cdots.$$

与实值情况相同, $R(0)$ 是实的, 并且

$$R(0)\geqslant|R(u)|,\qquad u=0,\pm1,\pm2,\cdots.$$

但对称性现在是以埃尔密特形式出现：

$$R(u) = \overline{R(-u)},$$

并且半正定性是指对于任意正整数 n 及任意复数 $\alpha_1, \cdots, \alpha_n$,

$$\sum_{j=1}^{n}\sum_{k=1}^{n}\alpha_i\overline{\alpha}_j R(i-j) \geqslant 0.$$

下面定理是定理 7.1 的一般化. 其证明步骤与前述定理完全平行, 这里从略.

定理 7.3 设复值函数 $R(u), u = 0, \pm 1, \cdots$, 是已知的. 则下面命题等价：

(a) $R(u)$ 是具有零均值和单位方差复值协方差平稳过程的协方复函数.

(b) $R(0) = 1$, 对于 $u = 0, 1, \cdots, R(u) = \overline{R(-u)}$, 并且 $R(u)$ 是半正定的.

(c) 在 $[-\pi, \pi]$ 上存在一个概率分布函数 F(不一定对称), 使得

509

$$R(u) = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}u\omega}\mathrm{d}F(\omega).$$

9.8 高斯系统

在本节将构造一个具有对称谱分布函数 $F(\omega)(-\pi \leqslant \omega \leqslant \pi)$ 的实值高斯平稳过程 $\{X_n\}$. 我们利用下面公式 (记号 $Z^{(i)}(\mathrm{d}\omega)$ 与 $\mathrm{d}Z^{(i)}(\omega)$ 意义相同.)

$$X_n = \int_0^{\pi} \cos n\omega Z^{(1)}(\mathrm{d}\omega) + \int_0^{\pi} \sin n\omega Z^{(2)}(\mathrm{d}\omega), \tag{8.1}$$

其中 $\{Z^{(i)}(\omega); 0 \leqslant \omega \leqslant \pi\}, i = 1, 2$, 是高斯过程, 彼此独立, 且具有独立增量, 并且

$$E[\{Z^{(i)}(\omega_2) - Z^{(i)}(\omega_1)\}^2] = 2\{F(\omega_2) - F(\omega_1)\}, \quad i = 1, 2, 0 \leqslant w_1 \leqslant w_2 \leqslant \pi$$

一般来说, 对于每个协方差平稳过程都存在形如 (8.1) 的表达式, 但若 $\{X_n\}$ 不是高斯的, $Z^{(i)}$ 也不是高斯的. 下面我们仅介绍这方面一些简单的结果.

高斯系统

设 T 是一个抽象集合, 并且 $\{X(t); t \in T\}$ 是一随机过程. 我们称 $\{X(t), t \in T\}$ **是高斯系统或高斯过程**, 如果对每个 $n = 1, 2, \cdots$ 和 T 的每个有限子集 $\{t_1, \cdots, t_n\}$, 随机向量 $(X(t_1), \cdots, X(t_n))$ 具有多元正态分布. 等价地, 我们可要求对每个 n 和每个线性组合

$$\alpha_1 X(t_1) + \cdots + \alpha_n X(t_n), \quad \alpha_i 为实数,$$

都具有一维正态分布. 每个高斯系统由其均值和协方差函数唯一确定, 记其为

$$\mu(t) = E[X(t)], \qquad t \in T.$$

和

$$\Gamma(t_1, t_2) = E[\{X(t_1) - \mu(t_1)\}\{X(t_2) - \mu(t_2)\}], \qquad t_i \in T.$$

协方差函数是半正定的, 即对每个 $n=1,2,\cdots$, 实数 $\alpha_1,\cdots,\alpha_n$ 及 T 中元素 $t_1, t_2, \cdots, t_n$,

$$\sum_{i,j=1}^{n} \alpha_i \alpha_j \Gamma(t_i, t_j) \geqslant 0. \tag{8.2}$$

为验证这点我们只要计算 $\sum_{i=1}^{n} \alpha_i \{X(t_i) - \mu(t_i)\}$ 的方差. 事实上,

510

$$\begin{aligned} 0 &\leqslant E\left[\left\{\sum_{i=1}^{n} \alpha_i (X(t_i) - \mu(t_i))\right\}^2\right] \\ &= \sum_{i,j=1}^{n} \alpha_i \alpha_j E\left[\{X(t_i) - \mu(t_i)\}\{X(t_j) - \mu(t_j)\}\right] \\ &= \sum_{i,j=1}^{n} \alpha_i \alpha_j \Gamma(t_i, t_j). \end{aligned}$$

反之, 任意给定均值函数 $\mu(t)$ 和半正定协方差函数 $\Gamma(t_1, t_2)$, 存在对应的高斯系统. 为说明这一事实, 只要注意到每个有限族 $\{X(t_1), \cdots, X(t_n)\}, t_i \in T$, 都对应于一个具有均值向量 $\{\mu(t_1), \cdots, \mu(t_n)\}$ 和协方差矩阵 $\|\Gamma(t_i, t_j)\|_{i,j=1}^{n} (t_i \in \Gamma)$ 的多元正态分布. 这就确定了过程 $\{X(t); t \in T\}$ 的相容的有限维分布族. 再根据 1.4 节即知所述的高斯系统必存在.

高斯随机测度

设 E 是有限维空间的一个子集, $g(x)$ 是定义在 E 上非负函数, 并满足 $\int_E g(x)\mathrm{d}x < \infty$. 对 E 的每个子集 A, 定义

$$m(A) = \int_A g(x)\mathrm{d}x.$$

我们断定表达式

$$\Gamma(A, B) = m(A \cap B)$$

定义一个指标参数是 E 的子集的半正定函数. 几乎和前面完全相同, 我们计算

$$0 \leqslant \int_E \left\{\sum_{i=1}^{n} \alpha_i I_{A_i}(x)\right\}^2 g(x)\mathrm{d}x$$

$$= \sum_{i,j=1}^{n} \alpha_i \alpha_j \int_E I_{A_i}(x) I_{A_j}(x) g(x) \mathrm{d}x \tag{8.3}$$

$$= \sum_{i,j=1}^{n} \alpha_i \alpha_j m(A_i \cap A_j),$$

其中

511

$$I_A(x) = \begin{cases} 1, & \text{如果} \quad x \in A, \\ 0, & \text{如果} \quad x \notin A. \end{cases}$$

一个具有均值为 0, 协方差为 $\Gamma(A, B) = m(A \cap B), A, B \subset E$的高斯系统 $\{Z(A), A \subset E\}$ 称为由 m 导出的**高斯随机测度**. 倘若 $\int_E g(x)\mathrm{d}x = \infty$, 我们仍然可定义 $Z(A)$, 但不对 E 的所有子集, 而仅对使得 $\int_A g(x)\mathrm{d}x < \infty$ 的那些子集 A.

假设 E 是某个有限维空间, 并对所有 $x, g(x) \equiv 1$. 这时, 对应的高斯随机测度 $\{Z(A)\}$ 是以 E 的子集为指标的平稳过程. 由 g 的恒定性保证 $(Z(A_1+x), \cdots, Z(A_n+x))$ 和 $(Z(A_1), \cdots, Z(A_n))$ 具有相同的分布, 其中 $A_1 + x$ 是集合 A_1 的 x 平移, 即 $A_1 + x = \{y : y = x + z, z \in A_1\}$.

假设 $E = [a, b]$, 其中 $0 \leqslant a \leqslant b < \infty$, 并设 $G(x)$ 是 $[a, b]$ 上有界不减函数. 当 $I = (x_1, x_2]$ 时 (其中 $a \leqslant x_1 < x_2 \leqslant b$) 定义

$$G(I) = G(x_2) - G(x_1),$$

则 G 是半开半闭区间 $I = (x_1, x_2] \subset E$ 的非复函数, 置 $\Gamma(I_1, I_2) = G(I_1 \cap I_2)$, 可证明 $\Gamma(I_1, I_2)$ 是半正定的. 当 G 具有导数 g 时, 证明完全与 (8.3) 相同. 至于一般情况, 方法同样, 只要用增量 $\mathrm{d}G(x)$ 代替 $g(x)\mathrm{d}x$, 并利用第 1 章里所叙述的有关黎曼 - 斯蒂尔切斯积分的性质即可得出结论.

设 $Z(I)$ 是区间 $I = (x, y], a \leqslant x < y \leqslant b$ 上的高斯随机测度, 并满足 $E[Z(I)^2] = G(I)$. 我们定义关于连续函数 $f(x)$ 的积分

$$\int_a^b f(x) Z(\mathrm{d}x)$$

为和

$$\sum_{i=0}^{n-1} f(x_i) Z(I_i)$$

的均方极限, 其中 $I_i = (x_i, x_{i+1}], a = x_0 < x_1 < \cdots < x_n = b$.

设 $\mathscr{P}=\{x_i\}, a=x_0<x_1<\cdots<x_n<b$, 确定 $[a,b]$ 的一个分割. 我们称分割 $\mathscr{P}'=\{x_i'\}$ 是分割 $\mathscr{P}$ 的加细, 如果对任意 $x_i\in\mathscr{P}$, 必有 $x_i\in\mathscr{P}'$. 令

$$\oint(f;\mathscr{P})=\sum_{i=0}^{n-1}f(x_i)Z(I_i),\quad I_i=(x_i,x_{i+1}].$$

若 $\oint(f;\mathscr{P})$ 是非随机的, 我们可通过不断加细分割, 并使得和 $\oint(f;\mathscr{P})$ 收敛来定义一个积分. 此处 $\oint(f;\mathscr{P})$ 虽是随机的, 我们仍可仿效相同方法, 只不过要用均方收敛来代替普通的实数收敛. 准确地说, 我们将证明对于任给的 $\varepsilon>0$, 存在一个分割 $\mathscr{P}$, 使得对所有加细分割 $\mathscr{P}'$, 满足

512

$$\left\|\oint(f;\mathscr{P})-\oint(f;\mathscr{P}')\right\|<\varepsilon,$$

其中 $\|\cdot\|$ 是 9.2 节中的平均平方距离. 利用该节所说的平均平方距离的**完备性**, 可推知当分割不断加细时, 存在一个在均方意义下的极限. 这个极限就定义为积分

$$\oint(f)=\int_a^b f(x)Z(\mathrm{d}x).$$

由于 $f(x)$ 在 $[a,b]$ 上连续必定一致连续, 我们可选择分割 $\mathscr{P}=\{x_i\}$, 使函数 f 在任何子区间 $I_i=(x_i,x_{i+1}]$ 上的振幅小于

$$\delta=\varepsilon/\sqrt{\{G(b)-G(a)\}},$$

其中 $\varepsilon>0$ 是任意预先给定的. 设 $\mathscr{P}'=\{x_j\}$ 是 $\mathscr{P}$ 的加细. 每个区间 $I_j'=(x_j',x_{j+1}']$ 必包含于某个 $I_i=(x_i,x_{i+1}]$ 当然 $\varepsilon_j=f(x_i)-f(x_j')<\delta$. 于是

$$\begin{aligned}\left\|\oint(f;\mathscr{P})-\oint(f;\mathscr{P}')\right\|^2&=E\left[\left\{\oint(f;\mathscr{P})-\oint(f;\mathscr{P}')\right\}^2\right]\\&=E\left[\left\{\sum_{i=0}^{n-1}f(x_i)Z(I_i)-\sum_{j=0}^{m-1}f(x_j')Z(I_j')\right\}^2\right]\\&=E\left[\left\{\sum_{j=0}^{m-1}\varepsilon_jZ(I_j')\right\}^2\right]\\&=\sum_{i,j=0}^{m-1}\varepsilon_i\varepsilon_jG(I_j'\cap I_j')\\&\leqslant\delta^2\sum_{i=0}^{m-1}G(I_j')=\delta^2\{G(b)-G(a)\}=\varepsilon^2.\end{aligned}$$

由于 ε 是任意的, 所以当分割无限加细时和在均方意义下收敛. 其极限就是积分

$$\oint(f) = \int_a^b f(x)Z(\mathrm{d}x).$$

513 作为均方收敛极限, 这是一个随机变量.

积分 $\oint(\cdot)$ 具有通常积分的大多数性质. 例如线性性质: 对任意实数 α, β 和连续函数 f_i,

$$\oint(\alpha f_1 + \beta f_2) = \alpha \oint(f_1) + \beta(f_2),$$

只要我们把上面等号 "=" 理解为在均方距离意义下的相等, 即

$$\left\| \oint(\alpha f_1 + \beta f_2) - \alpha \oint(f_1) - \beta \oint(f_2) \right\| = 0.$$

由于 $\oint(f)$ 是随机变量, 我们可以计算它的期望和方差. 在 9.2 节中已指出期望和方差在均方收敛意义下是连续的. 这样,

$$E\left[\oint(f)\right] = \lim E\left[\sum_i f(x_i)Z(I_i)\right] = 0.$$

并且对任何连续函数 $f(x), h(x)$ 有

$$\begin{aligned}
E\left[\oint(f)\oint(h)\right] &= \lim E\left[\sum_i f(x_i)Z(I_i)\sum_j h(x_j)Z(I_j)\right] \\
&= \lim \sum_{i,j} f(x_i)h(x_j)G(I_i \cap I_j) \\
&= \lim \sum_{i,j} f(x_i)h(x_j) \int I_{I_i}(x) I_{I_j}(x) \mathrm{d}G(x) \\
&= \lim \int \left\{\sum_i f(x_i) I_{I_i}(x)\right\} \left\{\sum_j h(x_j) I_{I_j}(x)\right\} \mathrm{d}G(x) \\
&= \int_a^b f(x)h(x)\mathrm{d}G(x). \qquad (8.4)
\end{aligned}$$

看起来很简单, 但它都是关于高斯随机测度积分的最重要性质.

高斯平稳过程的谱表示

设 $Z^{(1)}$ 和 $Z^{(2)}$ 是 $[0,\pi]$ 上的独立高斯随机测度且满足

514

$$E[\{Z^{(i)}(I)\}^2] = 2F(I),$$

其中, $I=(\omega_1,\omega_2], 0\leqslant \omega_1-\omega_2\leqslant \pi$, 并且 $F(\omega), -\pi\leqslant\omega\leqslant\pi$, 是已知的对称分布函数. 我们假设 $F(0)=F(0-)$. 读者可能希望这样的随机测度可用下面公式表达

$$Z^{(i)}(I)=\sqrt{2}\{B(F(\omega_2))-B(F(\omega_1))\},$$

其中 $\{B(t);t\geqslant 0\}$ 是标准布朗运动 (见第 7 章). 利用随机积分 $\oint(\cdot)$, 定义

$$\begin{aligned}X_n &= \int_0^{\pi}\cos n\omega Z^{(1)}(\mathrm{d}\omega)+\int_0^{\pi}\sin n\omega Z^{(2)}(\mathrm{d}\omega), \quad n=0,\pm1,\cdots, \\ &= \oint_1(\cos n\omega)+\oint_2(\sin n\omega).\end{aligned} \tag{8.5}$$

则 $E[X_n]=0$, 并根据 (8.4) 有

$$\begin{aligned}&E[X_nX_{n+u}]\\ &=E\left[\left\{\oint_1(\cos n\omega)+\oint_2(\sin n\omega)\right\}\left\{\oint_1(\cos(n+u)\omega)+\oint_2(\sin(n+u)\omega)\right\}\right]\\ &=E\left[\oint_1(\cos n\omega)\oint_1(\cos(n+u)\omega)\right]+E\left[\oint_2(\sin n\omega)\oint_2(\sin(n+u)\omega)\right]\\ &=2\int_0^{\pi}\cos n\omega\cos(n+u)\omega\mathrm{d}F(\omega)+2\int_0^{\pi}\sin n\omega\sin(n+u)\omega\mathrm{d}F(\omega)\\ &=\int_{-\pi}^{\pi}\cos u\omega\mathrm{d}F(\omega).\end{aligned}$$

由定理 7.1 推得 $\{X_n\}$ 是具有对称谱分布函数 F 的协方差平稳过程.

若引进复值高斯随机测度, 则对复值高斯平稳过程有上面平行的结果.

等式 (8.5) 利用高斯随机测度 $Z^{(i)}$ 表示平稳高斯过程 $\{X_n\}$. 我们不加证明地叙述相反的命题. 设 $\{X_n\}$ 是具有谱分布函数 F 的高斯平稳过程. 公式

$$Z^{(1)}(\lambda)=\frac{1}{2\pi}\left\{\lambda X_0+\sum_{n=1}^{\infty}\frac{1}{n}(X_n-X_{-n})\sin\lambda n\right\}$$

和

$$Z^{(2)}(\lambda)=\frac{1}{2\pi}\sum_{n=1}^{\infty}\frac{1}{n}(X_n+X_{-n})\cos\lambda n,$$

确定了具有独立增量的相互独立的高斯过程, 其中无限求和号理解为在均方意义下. 515

如果 $I=(\omega_1,\omega_2]$ 和 $Z^{(i)}(I)=Z^{(i)}(\omega_2)-Z^{(i)}(\omega_1)$, 则

$$E[\{Z^{(i)}(I)\}^2]=F(\omega_2)-F(\omega_1).$$

9.9 平稳点过程

设 $\mathscr{A}$ 是非负半直线 $[0,\infty)$ 的子集族, 并假设形如 $(t,s], 0 \leqslant t < s$ 的每个区间是 $\mathscr{A}$ 的成员. 点过程 $\{N(A); A \in \mathscr{A}\}$ 称为**平稳的**, 如果对于每个实数 h, 任意正整数 k, 以及每个区间组

$$(t_1, s_1], \cdots, (t_k, s_k],$$

k 维随机向量

$$(N(t_1, s_1], \cdots, N(t_k, s_k])$$

与

$$(N(t_1 + h, s_1 + h], \cdots, N(t_k + h, s_k + h])$$

具有相同的联合分布. 当然, 若 $\mathscr{A}$ 是整条直线的子集族, 或有限维欧几里的空间的子集族, 也可类似地定义平稳点过程.

如果 $\{N(A), A \in \mathscr{A}\}$ 是平稳点过程. 则对每个固定 $t, s \geqslant 0$, 非负整数值随机过程

$$W(h) = N(t + h, s + h], \quad h \geqslant 0,$$

是平稳过程, 因而也就具有本章前 8 节所叙述的结果. 然而, 点过程的本身特性须特别加以研究.

假设 $\{X(t); t \geqslant 0\}$ 是平稳过程且它的每一条轨道 $X(t)$ 是 t 的连续函数. 固定水平 u 并令 $N_u(s,t]$ 是轨道 $X(t)$ 在时间区间 $(s,t]$ 穿过 u 的次数. 则 $\{N_u(s,t]; 0 \leqslant s < t < \infty\}$ 是平稳点过程. 在 9.10 节中, 对于某些高斯平稳过程 $\{X(t)\}$, 我们将要计算均值 $E[N_u(0,t]]$. 在本节我们只说明一个平稳更新过程 (5.7 节) 如何导致一个平稳点过程.

516

设 F 是非负随机变量 X 的分布函数, X 具有有限均值

$$\mu = E[X] = \int_0^\infty x \mathrm{d}F(x)$$

$$= \int_0^\infty [1 - F(y)] \mathrm{d}y,$$

后面的等式可由分部积分得到 (也可参考第 1 章问题 9). 这样,

$$g(y) = \mu^{-1}[1 - F(y)], \qquad y \geqslant 0,$$

是非负的并且积分为 1. 因此它是一个概率密度函数.

现在令 $X_0, X_1, X_2, \cdots$ 是独立随机变量, 其中 X_0 具有概率密度函数 g, 而当 $i \geqslant 1$ 时, X_i 具有分布函数 F. 在每个时刻于 $s_n = X_0 + \cdots + X_n (n \geqslant 0)$ 处安置一个 "点", 这样就构造了一个点过程. 准确地说, 对任何区间 $A = (t, s]$, 令 $N(A)$ 是满足 $t < S_n \leqslant s$ 的指标 n 的数目.

我们断言以这种方式构造的点过程是平稳点过程. 此时具体地表达向量 $\boldsymbol{N}(0) = (N(t_1, s_1], \cdots, N(t_k, s_k])$ 的联合分布函数是非常困难的. 但我们不需要确定这个分布, 仅需要证明它与位移向量 $\boldsymbol{N}(h) = (N(t_1 + h, s_1 + h], \cdots, N(t_k + h, s_k + h])$ 的分布是相同的. 记 $\boldsymbol{X}$ 为无限维向量 $(X_0, X_1, \cdots)$, 则 $\boldsymbol{N}(0)$ 是 $\boldsymbol{X}$ 的向量值函数. 令 f 表示这个函数. 并令 f_n 是把 $\boldsymbol{X}$ 映射为 $\boldsymbol{N}(h)$ 的向量值函数. 于是, 我们想要证明 $\boldsymbol{N}(0) = f(\boldsymbol{X})$ 和 $\boldsymbol{N}(h) = f_h(\boldsymbol{X})$ 具有相同的分布. 但是采用这种方法, 问题将变得十分困难.

证明的技巧在于把 $\boldsymbol{N}(h)$ 表示为与 $\boldsymbol{N}(0)$ 相同的形式, 只不过把函数 f 作用在位移序列 $\boldsymbol{X}' = (X_0', X_1', \cdots)$ 上. 我们断言

$$\boldsymbol{N}(h) = f(\boldsymbol{X}') \tag{9.1}$$

其中

$$X_0' = S_M - h, \quad X_k' = X_{M+k} \quad k \geqslant 1 \tag{9.2}$$

并且 M 由

$$S_{M-1} < h \leqslant S_M \tag{9.3}$$

517

确定. 因此, 为证明 $\boldsymbol{N}(h) = f(\boldsymbol{X}')$ 与 $\boldsymbol{N}(0) = f(\boldsymbol{X})$ 具有相同的分布, 我们仅需要证明无限维向量 $\boldsymbol{X}' = (X_0', X_1', \cdots)$ 和无限维向量 $\boldsymbol{X} = (X_0, X_1, \cdots)$ 具有相同的分布, 因为函数 f 对于 $\boldsymbol{N}(0)$ 和 $\boldsymbol{N}(h)$ 都是相同的.

验证等式 (9.1) 并不太难. 我们仅考虑一维情形, 这时

$$\boldsymbol{N}(0) = N(t, s],$$

其中 $t < s$ 是固定的. 令 $\#\{\}$ 表示括号中的数目, 则

$$\begin{aligned} N(t, s] &= \#\{n : t < s_n \leqslant s\} \\ &= f(\boldsymbol{X}) \end{aligned}$$

则

$$\begin{aligned} N(t + h, s + h] &= \#\{n \geqslant 0 : t + h < S_n \leqslant s + h\} \\ &= \#\{n \geqslant M : t + h < S_n \leqslant s + h\} \\ &= \#\{n \geqslant M : t < S_n - h \leqslant s\} \end{aligned}$$

$$= \#\{m \geqslant 0 : t < S_{m+M} - h \leqslant s\}$$

$$= \#\{m \geqslant 0 : t < S'_m \leqslant s\}$$

$$= f(\boldsymbol{X}')$$

其中, 如前所述, X' 和 M 由等式 (9.2) 和 (9.3) 给出, 并且 $S'_m = X'_0 + \cdots + X'_m$.

这样, 为验证 $\{N(A); A \in \mathscr{A}\}$ 是平稳点过程, 只需证明

$$\boldsymbol{X}' = (S_M - h, X_{M+1}, X_{M+2}, \cdots)$$

和

$$\boldsymbol{X} = (X_0, X_1, X_2, \cdots)$$

具有相同的分布. 由于 M 由随机变量 $X_0, \cdots, X_M$ 所确定, 因此与 $X_{M+1}, X_{M+2}, \cdots$ 独立. 由此推得 X_{M+k} 的分布与已知 $M = n$ 条件下 X_{M+k} 的条件分布相同. 利用独立性, 它又与 X_{n+k} 固而与 X_k 的分布相同. 类似地, $X_{M+1}, X_{M+2}, \cdots$ 是独立的, 并且皆与 $S_M - h$ 独立. 这样 $X'_1, X'_2, \cdots$ 是独立同分布的, 具有分布函数 F, 并且独立于 $X'_0 = S_M - h$. 现在剩下的只需要证明 $X'_0 = S_M - h$ 具有概率密度函数 $g(y) = \mu^{-1}[1 - F(y)], y \geqslant 0$. 但这一点已在 5.7 节证明了, 其中 $X'_0 = S_M - h$ 相当于在时刻 h 的剩余寿命. 这样, 每个平稳更新过程导出了在非负半直线上的一个平稳点过程.

518

下面定理把上述结果推广到整条直线上.

定理 9.1 设 $\{X_n; n = 0, \pm 1, \pm 2, \cdots\}$ 是独立的正随机变量序列. 我们假设 $X_k (k = \pm 1, \pm 2, \cdots)$ 具有共同的概率密度函数 $f(x) (x \geqslant 0)$. 并设 X_0 具有和 $X_0^+ + X_0^-$ 的分布, 其中 X_0^+, X_0^- 的联合分布由密度函数

$$\mu^{-1} f(x^+ + x^-), \quad 对于 x^+ \geqslant 0, x^- \geqslant 0$$

确定. 如果 "点" 安放在实直线上 $S_n = X_0^+ + \cdots + X_n$ 和 $S_{-n} = -X_0^- - \cdots - X_{-n}$ 处, $n = 0, 1, 2, \cdots$, 则所导出的点过程是平稳的.

9.10 水平交叉问题

在本节我们假设 $\{X(t); -\infty < t < \infty\}$ 是零均值高斯平稳过程, 并且其每一条轨道都是 t 的连续函数. 我们固定水平 a 并考虑轨道 $X(t)$ 在时间区间 $(0, T]$ 内穿过 a 的次数. 这个量在通讯理论以及其他许多领域中都是非常重要的. 要计算这个随机变量的分布是非常困难的. 实际上, 在很多情况下显式表达式是未知的, 我们仅

局限于计算其一阶矩, 或均值, 并且还附加下面条件:

$$\lambda_2 = -\frac{\mathrm{d}^2}{\mathrm{d}t^2}R(t)\Big|_{t=0} < \infty, \tag{10.1}$$

其中 $R(t)$ 是过程的协方差函数.

我们把下面推导中用到的一般结论列为引理.

对点过程使用记号

$$N(I) = N(s,t], \quad I = (s,t], \quad 0 \leqslant s < t$$

我们说 $\{N(I)\}$ 没有多重事件, 如果

$$\lim_{n\to\infty} N(t-(1/n),t] = 0 \text{ 或 } 1, \text{ 对于所有} t. \tag{10.2}$$

9.9 节的更新点过程就提供了具有这样性质的一个例子.

519

引理 10.1 设 $\{N(I)\}$ 是没有多重事件的点过程, 固定 $T > 0$, 并分割区间 $(0,T]$ 为 n 个子区间

$$I_{ni} = ((i-1)T/n, iT/n], i = 1,\cdots,n;\ n = 1,2,\cdots.$$

则

$$E[N(0,T]] = \lim_{n\to\infty}\sum_{i=1}^{n}\Pr\{N(I_{ni}) \geqslant 1\}. \tag{10.3}$$

证明 记

$$\chi_{ni} = \begin{cases} 1, & 若 N(I_{ni}) \geqslant 1. \\ 0, & 其他. \end{cases}$$

及 $N_n = \sum\limits_{i=1}^{n}\chi_{ni}$. 则 $N_n \leqslant N_{n+1} \leqslant N(0,T]$, 并由 (10.2) 知, $\lim\limits_{n\to\infty} N_n = N(0,T]$. 可以验证极限号与期望可以交换, 由此得到

$$\lim_{n\to\infty}\sum_{i=1}^{h}\Pr\{N(I_{ni}) \geqslant 1\} = \lim_{n\to\infty} E[N_n] = E[N(0,T)].$$

■

等式 (10.3) 表明可使用小区间上事件的分布来表示在 $[0,T]$ 上事件的平均数. 我们感兴趣的是由连续过程穿过水平 a 所确定的事件. 下一步我们来研究过程 $X(t)$ 在一个区间里穿过水平 a 的次数. 为此, 我们需要关于"穿过"的恰当定义. 固定 a, 称 $X(t)$ 在 t_0 处穿过 a, 如果对于任给正数 ε, 存在点 t_1 和 t_2, 一方面满足

$$|t_i - t_0| < \varepsilon, \quad i = 1,2,$$

另一方面

$$[X(t_1)-a]\cdot[X(t_2)-a]<0.$$

注意, 若在 t_0 足够小的邻域, $X(t)$ 不超过 (或不小于)a 但在 t_0 处达到 a, 这时不能看作穿过 a. 令 $N_a(0,T]$ 是 $X(t)$ 在区间 $(0,T]$ 穿过 a 的数目.

引理 10.2 设 $\{X(t)\}$ 是每条样本轨道都连续的随机过程, 并且对每个固定 $t, \Pr\{X(t)=0\}=0$ 则

520

$$E[N_a(0,T]]=\lim_{n\to\infty}\left\{\sum_{i=1}^{n}\Pr\left\{X\left(\frac{(i-1)T}{n}\right)<a<X\left(\frac{iT}{n}\right)\right\}\right.$$
$$\left.+\sum_{i=1}^{n}\Pr\left\{X\left(\frac{(i-1)T}{n}\right)>a>X\left(\frac{iT}{n}\right)\right\}\right\}.$$

证明 这个结果差不多与引理 10.1 是相同的, 令

$$\chi'_{ni}=\begin{cases}1, & \text{若 } X\left(\frac{(i-1)T}{n}\right)<a<X\left(\frac{iT}{n}\right) \text{ 或 } X\left(\frac{(i-1)T}{n}\right)>a>X\left(\frac{iT}{n}\right)\\ 0, & \text{其他.}\end{cases}$$

显然, $N'_n=\sum_{i=1}^{n}\chi'_{ni}\leqslant N_n$, 所以由引理 10.1 有

$$\limsup_{n\to\infty}E[N'_n]\leqslant E[N_a(0,T]],$$

另一方面, 对每个固定的 n, 当 m 足够大时, 显然有 $N_n\leqslant N'_m$, 因为在这种情况下, 例如, $(X(t_1)-a)<0,(X(t_2)-a)>0$, 对满足

$$\frac{(i-1)T}{n}<t_1<t_2<\frac{iT}{n}$$

的 t_1,t_2 成立, 将推得

$$X\left(\frac{(j-1)T}{m}\right)<a<X\left(\frac{jT}{m}\right),$$

对某个 I_{ni} 的子区间 I_{mj} 成立. [我们已排除 $X(t)$ 在形如 $t=rT$ 的点上穿过 a 这个零概率事件, 其中 r 是 $(0,1]$ 上有理数.] 这样,

$$\liminf_{m\to\infty}E[N'_m]\geqslant E[N_n],$$

并且

$$\liminf_{m\to\infty}E[N'_m]\geqslant\liminf_{n\to\infty}E[N_n]=E[N_a(0,T)].$$

这就完成了证明. ■

因此, 在一个区间上穿过 a 的平均数的计算可以通过研究下面比较简单的事件.

$$\left\{X\left(\frac{(i-1)T}{n}\right) < a < X\left(\frac{iT}{n}\right)\right\}$$

和

$$\left\{X\left(\frac{(i-1)T}{n}\right) > a > X\left(\frac{iT}{n}\right)\right\}$$

521

来实现. 如果过程是平稳的, 这些事件分别与事件

$$A(n) = \{X(0) < a < X(T/n)\}$$

和

$$B(n) = \{X(0) > a > X(T/n)\}$$

具有相同的概率.

定理 10.1 令 $\{X(t); t \geqslant 0\}$ 是零均值高斯平稳过程, 并且其每条轨道 $X(t)$ 是 t 的连续函数. 假设协方差函数 $R(t)$ 满足 (10.1), 并设 $X(t)$ 的方差 $\sigma^2 = R(0)$. 则在 $(0,T]$ 上穿过水平 a 的平均次数可由下式计算

$$E[N_a(0,T]] = \frac{T}{\pi}(\lambda_2/\sigma^2)^{1/2}\exp(-a^2/2\sigma^2).$$

证明 由平稳性可知, 最后结果一定与 T 成正比, 因此我们仅需要处理 $T=1$ 的情况. 根据前面的引理. 我们的任务是计算

$$\begin{aligned} E[N_a(0,1]] &= \lim_{n\to\infty} n[\Pr\{A(n)\} + \Pr\{B(n)\}] \\ &= \lim_{n\to\infty} 2^n[\Pr\{A(2^n)\} + \Pr\{B(2^n)\}]. \end{aligned}$$

由于 $A(2^n)$ 是由 $X(0)$ 和 $X(2^{-n})$ 确定的. 它们的联合分布是均值为 0, 方差为 $\sigma^2 = R(0)$ 的正态分布, 其相关系数为 $\rho(2^n) = R(2^{-n})/R(0)$.(关于二元正态分布的性质可见第 1 章.) 令

$$\zeta_n = 2^n[X(2^{-n}) - X(0)].$$

随机变量对 $X(0), \zeta_n$ 具有二元正态分布, 直接计算可以说明其均值为 0, 其协方差矩阵是

$$\left\|\begin{matrix} R(0) & -2^n[R(0)-R(2^{-n})] \\ -2^n[R(0)-R(2^{-n})] & 2^n\{2^n[R(0)-R(2^{-n})] - 2^n[R(-2^{-n})-R(0)]\} \end{matrix}\right\|$$

在计算上面矩阵元素时, 注意利用对称性 $R(-2^n) = R(2^{-n})$. 依假设 (10.1), 这个矩阵当 $n \to \infty$ 时收敛于

$$\left\| \begin{matrix} R(0) & R'(0) \\ R'(0) & R''(0) \end{matrix} \right\| = \left\| \begin{matrix} \sigma^2 & 0 \\ 0 & \lambda_2 \end{matrix} \right\| \tag{10.4}$$

522 其中 $R'(t)$ 和 $R''(t)$ 分别是 $R(t)$ 的一阶和二阶导数. 由 $R(t)$ 的对称性以及 $R''(0) < \infty$ 存在性的假定, 推得 $R'(0) = 0$. 令 $p_n(x, z)$ 表示 $X_{(0)}$ 和 ζ_n 的二元正态分布的密度. 注意 $B(2^n)$ 可以表示为

$$\begin{aligned} B(2^n) &= \{X(0) > a > X(2^{-n})\} \\ &= \{a < X(0) < a - 2^{-n}\zeta_n\}. \end{aligned}$$

计算

$$\begin{aligned} 2^n \Pr\{B(2^n)\} &= 2^n \Pr\{a < X(0) < a - 2^{-n}\zeta_n\} \\ &= 2^n \int_{-\infty}^{0} \int_{a}^{a-2^{-n}z} p_n(x, z) \mathrm{d}x\mathrm{d}z \\ &= \int_{-\infty}^{0} \int_{0}^{-z} p_n(a + 2^{-n}x, z) \mathrm{d}x\mathrm{d}z. \end{aligned}$$

利用第 1 章给出的二元正态分布的密度函数表示式, 并注意到 $X(0)$ 和 ζ_n 的协方差矩阵收敛于 (10.4) 的矩阵, 我们推得

$$\begin{aligned} \lim_{n\to\infty} p_n(a + 2^{-n}x, Z) &= \frac{1}{2\pi\sigma\sqrt{\lambda_2}} \exp\left\{-\frac{1}{2}\left[\left(\frac{a}{\sigma}\right)^2 + \frac{z^2}{\lambda_2}\right]\right\} \\ &= \frac{1}{\sigma}\phi\left(\frac{a}{\sigma}\right) \cdot \frac{1}{\lambda_2^{1/2}} \cdot \phi\left(\frac{z}{\lambda_2^{1/2}},\right) \end{aligned}$$

其中

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$

是标准正态分布密度.

这样,

$$\begin{aligned} \lim_{n\to\infty} 2^n \Pr\{B(2^n)\} &= \int_{-\infty}^{0} \int_{0}^{-z} \frac{1}{\sigma}\phi\left(\frac{a}{\sigma}\right) \frac{1}{\sqrt{\lambda_2}}\phi\left(\frac{z}{\sqrt{\lambda_2}}\right) \mathrm{d}x\mathrm{d}z \\ &= \frac{1}{\sigma}\phi\left(\frac{a}{\sigma}\right) \int_{0}^{\infty} \frac{z}{\sqrt{\lambda_2}}\phi\left(\frac{z}{\sqrt{\lambda_2}}\right) \mathrm{d}z \\ &= \frac{\sqrt{\lambda_2}}{\sigma}\phi\left(\frac{a}{\sigma}\right) \int_{0}^{\infty} y\phi(y) \mathrm{d}y \end{aligned}$$

$$= \frac{\sqrt{\lambda_2}}{\sigma}\phi\left(\frac{a}{\sigma}\right)\frac{1}{\sqrt{2\pi}}$$

$$= \frac{\sqrt{\lambda_2}}{\sigma 2\pi}\exp\left\{-\frac{1}{2}\left(\frac{a}{\sigma}\right)^2\right\}.$$

当计算 $\lim\limits_{n\to\infty} 2^n\Pr\{A(2^n)\}$ 时, 也得到相同的结果.

523

因此,

$$E[N_a(0,1)] = 2\lim_{n\to\infty} 2^n\Pr\{B(2^n)\}$$
$$= \frac{\sqrt{\lambda_2}}{\pi\sigma}\exp\left\{-\frac{1}{2}\left(\frac{a}{\sigma}\right)^2\right\},$$

这就是我们要证明的. ■

更细微地, 我们说 $X(t)$ 在 t_0 上穿水平线 a, 如果存在某个 $\varepsilon > 0$, 对于 $t_0 - \varepsilon < t < t_0, X(t) < a$, 对于 $t_0 < t < t_0 + \varepsilon, X(t) > a$. 令 $U_a(0,T]$ 是在 $(0,T]$ 内上穿的次数, 则

$$E[U_a(0,1]] = \lim_{n\to a} 2^n\Pr\{A(2^n)\} = \frac{\sqrt{\lambda_2}}{2\pi\sigma}\exp(-a^2/2\sigma^2)$$

在结束本节和本章之际, 我们在这个迷人的概率论领域中叙述一个比较近代的结果. 我们目的是激起读者进一步阅读有关内容的兴趣. 当水平 a 增加时, 上穿变成越来越少, 这样就出现了类似泊松性质的可能性. 自然, 当 a 增加时, $U_a(0,T]$ 将变得比较小, 从而我们需要正规化. 令

$$f(a) = \frac{\sqrt{\lambda_2}}{2\pi\sigma}\exp(-a^2/2\sigma^2),$$

和

$$N_a(t) = U_a(0, t/f(a)).$$

这时 $E[N_a(t)] = t$. 以下是这方面的结果:

定理 10.2 利用上面使用的记号. 令 $X(t)$ 是高斯平稳过程, 具有连续轨道和协方差函数 $R(t)$, 这里 $\lambda_2 = R''(0) < \infty$. 假设下面两个条件之一成立: (i) 当 $t \to \infty$ 时, $R(t)\ln t \to 0$; (ii) $\int_0^\infty R(s)^2 \mathrm{d}s < \infty$. 则 $\{N_a(t); t \geqslant 0\}$ 的分布当 $a \to \infty$ 时收敛于泊松过程的分布.

猜测为什么会有这样的结果, 我们注意到 (i) 和 (ii) 都蕴涵着渐近独立, 由此体现出了泊松过程的独立增量性质.

初 等 问 题

问题 1~5 均在下述条件之下考虑: 设 $\{X_n\}$ 和 $\{Y_n\}$ 是分别具有协方差函数 $R_X(u)$

524 和 $R_Y(u)$ 的零均值协方差平稳过程. 令 $R_{XY}(u)=E[X_nY_{n+u}], u=0,\pm1,\cdots$, 为互协方差函数. 并设 $\{\xi_n\}$ 是协方差函数为 $R_\xi(u)$ 的零均值协方差平稳过程, 且与 $\{X_n\},\{Y_n\}$ 是无关的. "最佳"预测量、估计量等都在最小均方误差的意义下.

1. (a) 求 X_{n+1} 的形如 $\widehat{X}^{(1)}_{(n+1)}=aX_n$ 的最佳预测量, 其中 a 是待定常数.

解答: $a^*=R_X(1)/R_X(0)$.

(b) 求 X_{n+1} 的形如 $\widehat{X}^{(2)}_{n+1}=aX_n+bX_{n-1}$ 的最佳预测量, 其中 a,b 是常数.

解答: $a^*=\dfrac{1}{\Delta}[R_X(1)R_X(0)-R_X(1)R_X(2)]$,

$$b^*=\frac{1}{\Delta}[R_X(0)R_X(2)-R_X(1)^2],$$

其中 $\Delta=R_X(0)^2-R_X(1)^2$.

(c) 试用 $R_X(u)$ 表达 $\widehat{X}^{(2)}_{n+1}$ 对 $\widehat{X}^{(1)}_{n+1}$ 在均方预测误差意义下的改进.

解答: $\dfrac{1}{R_X(0)}\{R_X(2)-\dfrac{R_X(1)^2}{R_X(0)}\}^2$.

2. (a) 试求 X_n 的形如 $\widehat{X}^{(1)}_n=aY_n$ 的最佳估计, 其中 a 是待定常数.

解答: $a^*=R_{XY}(0)/R_Y(0)$.

(b) 试求 X_n 的形如 $\widehat{X}^{(2)}_n=aYn+bY_{n-1}$ 的最佳估计, 其中 a,b 是常数.

解答: $a^*=\dfrac{1}{\Delta}[R_{XY}(0)R_Y(0)-R_{XY}(1)R_Y(1)]$,

525 $$b^*=\frac{1}{\Delta}[R_Y(0)R_{XY}(1)-R_Y(1)R_{XY}(0)],\text{ 其中 }\Delta=R_Y(0)^2-R_Y(1)^2.$$

3. 把 X_n 解释为信号, ξ_n 为噪声. 记 $Z_n=X_n+\xi_n$. 试求 X_n 的形如 $\widehat{X}_n=aZ_n+bZ_{n-1}$ 的最佳估计, 其中 a 和 b 是待定常数.

解答: $a^*=\dfrac{1}{\Delta}[R_X(0)\{R_X(0)+R_\xi(0)\}-R_X(1)\{R_X(1)+R_\xi(1)\}]$,

$$b^*=\frac{1}{\Delta}[\{R_X(0)+R_\xi(0)\}R_X(1)-\{R_X(1)+R_\xi(1)\}R_X(0)],$$

其中 $\Delta=\{R_X(0)+R_\xi(0)\}^2-\{R_X(1)+R_\xi(1)\}^2$.

4. 求 X_{n+k} 的形如 $\widehat{X}_{n+k}=aX_n+bX_{n+N}$ 的最佳插值, 其中 $1\leqslant k\leqslant N$ 是固定的, 并且 a,b 是待定常数.

解答: $a^*=\dfrac{1}{\Delta}[R_X(k)R_X(0)-R_X(N)R_X(N-k)]$,

$$b^*=\frac{1}{\Delta}[R_X(0)R_X(N-k)-R_X(k)R_X(N)],$$

其中 $\Delta=R_X(0)^2-R_X(N)^2$.

5. 固定 $N\geqslant1$, 并记 $Z_n=\sum\limits_{k=0}^{N}X_{n+k}$. 求 Z_n 的形如下式的最佳估计量

$$\widehat{Z}_n=aX_n+bX_{n+N},$$

其中 a, b 是待定常数.

6. 对于 $n=1,2,\cdots$ 令 $X_n=\cos(nU)$, 其中 U 均匀分布于区间 $[-\pi,\pi]$. 试证明 $\{X_n\}$ 是协方差平稳但不是严格平稳过程.

提示：第一部分使用三角恒等式

$$\cos x\cos y=\frac{1}{2}[\cos(x+y)+\cos(x-y)].$$

由对称性求得

$$E[\cos uU]=\begin{cases}1 & 若\ u=0\\ 0 & 若\ u=1,2,\cdots.\end{cases}$$

对于第二部分, 利用同样方法确定三阶乘积矩 $E[X_nX_{n+u}X_{n+u+h}]$ 依赖于 n.

7. 设 $\{B(t);t\geqslant 0\}$ 是标准布朗运动. 试计算 $X(t)=B(t+1)-B(t), t\geqslant 0$ 的协方差函数, 并证明 $X(t)$ 是严格平稳过程.

526

解答：$R(u)=\begin{cases}1-|u|, & 当\ |u|\leqslant 1,\\ 0, & 当\ |u|>1.\end{cases}$

8. 计算 $X_n=\sqrt{2}A\sin(\omega n+U)$ 的协方差函数, 其中 A 是均值为 0 方差为 σ^2 的随机变量, U 是独立于 A 的在 $[0,2\pi)$ 上均匀分布的随机变量, ω 是满足 $0\leqslant\omega<2\pi$ 的常数.

9. 假设 $\{X_n;n=0,1,\cdots\}$ 是零均值平稳过程, 且又是高斯过程和马尔可夫过程. 证明其协方差函数必具有形式 $R(u)=\sigma^2\lambda^{|u|}$, 其中 λ 为满足 $|\lambda|\leqslant 1$ 的某个固定常数.

10. 求对应于协方差函数 $R(0)=1$ 和 $R(u)=\alpha\gamma^{|u|}, u=\pm1,\pm2,\cdots$ 的谱密度函数, 其中 $0<\alpha<1$ 和 $|\gamma|<1$.

提示：记 $R(u)=R_1(u)+R_2(u)$, 其中

$$R_1(u)=\begin{cases}1-\alpha, & u=0,\\ 0, & u\neq 0.\end{cases}$$

及 $R_2(u)=\alpha\gamma^{|u|}$ 对所有 u.

11. 设有协方差平稳过程 $\{X_n\}$, 若在已知全过去下 X_n 的最佳线性预测量为 $\widehat{X}_n=a_1X_{n-1}+a_2X_{n-2}$. 试求其谱密度函数.

提示：由

$$\xi_n=X_n-\widehat{X}_n=X_n-a_1X_{n-1}-a_2X_{n-2}$$

推得

$$\sigma_\xi^2 f_\xi(\lambda)=\sigma_x^2|1-a_1\mathrm{e}^{\mathrm{i}\lambda}-a_2\mathrm{e}^{2\mathrm{i}\lambda}|^2 f_X(\lambda),$$

但由于 $\widehat{X}_n$ 是最佳预测量, $\{\xi_n\}$ 是不相关的, 且 $f_\xi(\lambda)=\dfrac{1}{2\pi}, -\pi\leqslant\lambda\leqslant\pi$, 从而解出 f_X.

问 题

1. 设 $\{\xi_n\}$ 是独立同分布随机变量序列, 其均值为 0, 方差为 σ^2. 令 $\{a_n\}$ 是实数序列. 证明当 $\sum_{n=0}^{\infty} a_n^2 < \infty$ 时 $X = \sum_{n=0}^{\infty} a_n\xi_n$ 必均方收敛. [特别地, $\sum(1/n)\xi_n$ 收敛!]. 令 527 $\{\eta_n\}$ 是零均值协方差平稳过程, 其协方差函数为 $R(v)$. 证明若

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}|a_i a_j R(i-j)| < \infty,$$

则 $\sum_{n=0}^{\infty} a_n\eta_n$ 均方收敛.

2. 设 $X, X_1, X_2, \cdots$ 是具有有限二阶矩随机变量序列. 证明 $\lim_{n\to\infty}\|X_n - X\| = 0$, 当且仅当以下两个条件满足: (i) 对任意满足 $E[Y^2] < \infty$ 的随机变量 Y, $\lim_{n\to\infty} E[X_nY] = E[XY]$.(ii) $\lim_{n\to\infty}\|X_n\| = \|X\|$.

3. 设

$$W_n = \sum_{j=1}^{q}\sigma_j\sqrt{2}\cos(\lambda_j n - V_j),$$

其中 σ_j, λ_j 是正常数, $j = 1, \cdots, q$, 而 $V_1, \cdots, V_q$ 是相互独立地均匀分布于区间 $(0, 2\pi)$. 证明 $\{W_n\}$ 是协方差平稳过程并计算其协方差函数.

4. 考虑满足下式的协方差平稳过程 $\{X_n\}$:

$$X_{n+1} = a_1X_n + a_2X_{n-1} + \xi_{n+1},$$

其中 a_1, a_2 为常数, $\{\xi_n\}$ 为零均值不相关随机变量列, 且 $E[\xi_n^2] = \sigma^2$ 及 $E[\xi_n X_{n-k}] = 0, k = 1, 2, \cdots$. 过程 $\{X_n\}$ 的相关函数记为 $\rho(u) = R(u)/R(0)$. 试证明 $\rho(u)$ 满足 Yule-Walker 等式

$$\rho(1) = a_1 + a_2\rho(1) \quad 和 \quad \rho(2) = a_1\rho(1) + a_2$$

并利用 $\rho(1)$ 和 $\rho(2)$ 确定 a_1 和 a_2.

5. 证明不存在一个协方差平稳过程 $\{X_n\}$ 满足随机差分方程 $X_n = X_{n-1} + \varepsilon_n$, 其中 $\{\varepsilon_n\}$ 是不相关随机变量序列, 其均值为 0, 共同方差 $\sigma^2 > 0$.

6. 设 $\{X_n\}_{n=-\infty}^{+\infty}$ 是零均值协方差平稳过程, 且具有协方差函数 $R(u) = r^{|u|}, u = 0, \pm 1, \cdots$, 其中 $|r| < 1$. 试求在已知全部过去 $X_n, X_{n-1}, \cdots$ 时, X_{n+1} 的最小均方线性预测量.

7. 设 $\{\varepsilon_n\}$ 是具有单位方差、零均值的不相关随机变量序列. $\{X_n\}$ 是如下形式的滑动平均过程

$$X_n=\varepsilon_n+\beta[\varepsilon_{n-1}+\gamma\varepsilon_{n-2}+\gamma^2\varepsilon_{n-3}+\cdots],$$

其中 β 和 γ 是常数, $|\gamma|<1,|\alpha|<1$ 且 $\alpha=\gamma-\beta$. 试求在已知全部过去 $X_n,X_{n-1},\cdots$ 的条件下 X_{n+1} 的最小均方线性预测量.

528

8. 设 $\{X_n\}$ 是协方差函数为 $R_X(u)$ 的零均值协方差平稳过程, 且其谱密度函数为 $f_X(\omega),-\pi\leqslant\omega\leqslant\pi$. 假设 $\{a_n\}$ 是满足条件 $\sum\limits_{i,j=0}^{\infty}|a_ia_jR(i-j)|<\infty$ 的实数列, 定义

$$Y_n=\sum_{k=0}^{\infty}a_kX_{n-k}.$$

证明 $\{Y_n\}$ 的谱密度函数由下式给出

$$\begin{aligned}f_Y(\omega)&=\frac{\sigma_X^2}{\sigma_Y^2}\left|\sum_{k=0}^{\infty}a_k\mathrm{e}^{\mathrm{i}k\omega}\right|^2f_X(\omega),\\&=\frac{\sigma_X^2}{\sigma_Y^2}\left[\sum_{j,k=0}^{\infty}a_ja_k\cos(j-k)\omega\right]f_X(\omega),\quad -\pi\leqslant\omega\leqslant\pi\end{aligned}$$

9. 试确定对应于协方差函数 $R(u)=\gamma^{|u|},u=0,\pm1,\cdots$,(此处 $|\gamma|<1$) 的谱密度函数.

答案: $f(\omega)=\dfrac{1-\gamma^2}{2\pi|1-\gamma\mathrm{e}^{\mathrm{i}\omega}|^2},-\pi<\omega<\pi.$

10. 考虑如下形式的自回归过程 $\{X_n\}$:

$$X_n=\beta_1X_{n-1}+\cdots+\beta_qX_{n-q}+\xi_n,$$

其中 $\{\xi_n\}$ 是零均值、单位方差的不相关随机变量序列, 并设 $x^q-\beta_1x^{q-1}-\cdots-\beta_q=0$ 的 q 个根绝对值均小于 1. 试求 $\{X_n\}$ 的谱密度函数.

答案: $f(\omega)=\left\{2\pi\sigma_X^2\left|1-\sum\limits_{k=1}^{q}\beta_k\mathrm{e}^{\mathrm{i}k\omega}\right|^2\right\}^{-1},\quad -\pi<\omega<\pi.$

11. 试计算滑动平均过程 $X_n=\xi_n+\alpha_1\xi_{n-1}$ 的谱密度函数, 其中 $\{\xi_n\}$ 是零均值、单位方差的不相关随机变量序列.

答案: $f(\lambda)=\dfrac{1+\alpha_1^2+2\alpha_1\cos\lambda}{2\pi(1+\alpha_1^2)}.$

12. 设 $\{X_n\}$ 是有限滑动平均过程

$$X_n=\sum_{r=0}^{q}\alpha_r\xi_{n-r},\qquad \alpha_0=1,$$

529

其中 $\alpha_0,\cdots,\alpha_q$ 是实的, $\{\xi_n\}$ 是具有单位方差、零均值的不相关随机变量序列. 试证明谱密度函数 $f(\lambda)$ 可以写为

$$f(\lambda)=\frac{1}{2\pi\sigma_X^2}\prod_{j=1}^{q}|\mathrm{e}^{\mathrm{i}\lambda}-z_j|^2.$$

其中 $z_1,\cdots,z_q$ 是

$$\sum_{r=0}^{q}a_r z^{q-r}=0$$

的 q 个根.

13. 试证明: 预测量

$$\widehat{X}_n=\alpha_1X_{n-1}+\cdots+\alpha_pX_{n-p}$$

在已知 $X_{n-1},\cdots,X_{n-p}$ 条件下 X_n 的所有线性预测量中是最佳的, 当且仅当

$$0=\int_{-\pi}^{\pi}\mathrm{e}^{\mathrm{i}k\lambda}\left[1-\sum_{l=1}^{p}\alpha_l\mathrm{e}^{-\mathrm{i}l\lambda}\right]\mathrm{d}F(\lambda),\quad k=1,\cdots,p,$$

其中 $F(\omega),-\pi\leqslant\omega\leqslant\pi$, 是协方差平稳过程 $\{X_n\}$ 的谱分布函数.

14. 设 $\{X_n\}$ 是零均值的协方差平稳过程, 具有正的谱密度函数 $f(\omega)$, 且方差 $\sigma_x^2=1$. 科尔莫戈罗夫公式指出

$$\sigma_e^2=\exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\ln 2\pi f(\omega)\mathrm{d}\omega\right\},$$

其中 $\sigma_e^2=\inf E[|\widehat{X}_n-X_n|^2]$ 是已知过去的条件下 X_n 的最小均方线性预测误差. 试对

$$R(u)=\gamma^{|u|},\quad u=0,\pm1,\cdots,|\gamma|<1,$$

验证科尔莫戈罗夫公式.

15. 试由

$$f(\lambda)=\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}R(k)\cos\lambda k,\quad -\pi\leqslant\lambda\leqslant\pi,$$

导出

$$R(u)=\int_{-\pi}^{\pi}(\cos\lambda u)f(\lambda)\mathrm{d}\lambda.$$

16. 考虑滑动平均过程

530

$$X_n=\sum_{k=0}^{\infty}a_k\xi_{n-k},$$

其中 $\{\xi_n\}$ 是零均值、单位方差的不相关随机变量序列, 且 $a_0,a_1,\cdots$ 是满足条件 $\sum a_k^2<\infty$ 的实数列. 试计算其协方差函数及谱密度函数.

17. 考虑下面非平稳线性模型

$$X_n = \theta_n + \varepsilon_n,$$

其中 $\theta_{n+1} = \theta_n + \xi_{n+1}, n = 0, 1, \cdots, \theta_0, \zeta_1, \zeta_2, \cdots$ 和 $\varepsilon_0, \varepsilon_1, \cdots$ 均为零均值不相关, $E[\theta_0^2] = v_0^2, E[\zeta_k^2] = v^2$, 和 $E[\varepsilon_k^2] = \sigma^2$, 此处 $v^2 = \alpha v_0^2, \alpha = v_0^2/(v_0^2 + \sigma^2)$.(我们把 $\{X_n\}$ 解释为对 θ 过程的噪音畸变观测.）试求在已知 $X_n, X_{n-1}, \cdots, X_0$ 时 X_{n+1} 的最小均方误差线性预测量.

答案:

$$\widehat{X}_0 = 0$$
$$\widehat{X}_k = \alpha X_{k-1} + (1-\alpha) X_{k-1}, \quad k = 1, 2, \cdots,$$

其中 $\alpha = v_0^2/(v_0^2 + \sigma^2)$.

18. 设 $\{X_k\}$ 是滑动平均过程

$$X_n = \sum_{j=0}^{\infty} \alpha_j \xi_{n-j}, \quad \alpha_0 = 1, \quad \sum_{j=0}^{\infty} \alpha_j^2 < \infty,$$

其中 $\{\xi_n\}$ 是零均值且具有共同方差 σ^2 的独立随机变量序列. 证明

$$U_n = \sum_{k=0}^{n} X_{k-1} \xi_k, \quad n = 0, 1, \cdots,$$

和

$$V_n = \sum_{k=0}^{n} X_k \xi_k - (n+1)\sigma^2, \quad n = 0, 1, \cdots,$$

关于 $\{\xi_n\}$ 是鞅.

19. 记 $\mathrm{i} = \sqrt{-1}$. 定义复值函数关于高斯随机测度的积分为实部和虚部积分之和. 类似地, 定义函数关于复值随机测度

$$\xi(I) = \xi(I) + \mathrm{i}\eta(I), \quad I = (s, t], \quad s < t,$$

的积分为实部和虚部之和. 由表达式

$$X_n = \int_0^{\pi} \cos n\omega Z^{(1)}(\mathrm{d}\omega) + \int_0^{\pi} \sin n\omega Z^{(2)}(\mathrm{d}\omega),$$

可得

$$X_n = \int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}n\omega} \zeta(\mathrm{d}\omega), \quad n = 0, \pm 1, \cdots,$$

置

$$\xi(s) = -\xi(-s) = \frac{1}{2}Z^{(1)}(s), \quad 0 \leqslant s \leqslant \pi,$$

和

$$\eta(s) = \eta(-s) = \frac{1}{2}Z^{(2)}(s), 0 \leqslant s \leqslant \pi.$$

注意到 $\xi(I) = \xi(-I)$ 和 $\eta(I) = -\eta(-I)$, 其中 $I = (s,t], -I = (-t,-s], 0 \leqslant s \leqslant t \leqslant \pi$. 试计算 $E[\zeta(I_1)\overline{\zeta(I_2)}]$, 其中 $I_i = (s_i, t_i]$, "——" 表示共轭复数.

20. 假设 X_0 具有概率密度函数

$$f(x) = \begin{cases} 2x, & 0 \leqslant x \leqslant 1, \\ 0, & \text{其他}. \end{cases}$$

若已知 $X_0, \cdots, X_n$, 则 X_{n+1} 是 $(1 - X_n, 1]$ 上均匀分布随机变量. 试证明 $\{X_n\}$ 是平稳遍历过程.

21. 证明每个独立同分布随机变量序列 $X_1, X_2, \cdots$ 构成一个遍历平稳过程.

22. 设 Z 是 $[0,1)$ 上均匀分布随机变量. 令 $X_0 = Z$ 和 $X_{n+1} = 2X_n(\mathrm{mod} 1)$, 即

$$X_{n+1} = \begin{cases} 2X_n, & \text{如果} \quad X_n < \dfrac{1}{2}, \\ 2X_n - 1, & \text{如果} \quad X_n \geqslant \dfrac{1}{2}. \end{cases}$$

(a) 证明如果 $Z = 0.Z_0Z_1Z_2\cdots$ 是 Z 的二进制展开: $Z = \sum_{k=0}^{\infty} 2^{-(k+1)}Z_k$, 则 $X_n = 0.Z_nZ_{n+1}\cdots$.(b) 证明 X_n 是平稳过程. (c) 证明 $\{X_n\}$ 是遍历的. (d) 利用遍历定理证明以概率 1,

$$\frac{1}{n}\sum_{k=0}^{n-1}\{2^kZ\} \to \frac{1}{2},$$

其中 $\{x\}$ 表示 x 的小数部分 ($\{x\} = x - [x]$, $[x]$ 是不超过 x 的最大整数).

23. 设 $\{\xi_n\}$ 是具有零均值单位方差的独立同分布随机变量序列. 证明: 每个滑动平均

532

$$X_n = \sum_{k=0}^{m} \alpha_k \xi_{n-k}, \quad n = 0, \pm 1, \cdots$$

是遍历的. 若 $\sum a_k^2 < \infty$, 同样的结论对于

$$Y_n = \sum_{k=0}^{\infty} a_k \xi_{n-k}$$

是否成立?

24. 一个随机过程 $\{X_n\}$ 称为是**弱混合的**，如果对于任意实数列 $(x_1, x_2, \cdots)$ 的集合 A, B,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\Pr\{(X_1, X_2, \cdots)\in A \text{ 和 } (X_k, X_{k+1}, \cdots)\in B\}$$
$$=\Pr\{(X_1, X_2, \cdots)\in A\}\times\Pr\{(X_1, X_2, \cdots)\in B\}.$$

证明每个弱混合过程是遍历的.

注　为了证明弱混合，只需证明对每个 $m=1,2,\cdots$，实向量 $(x_1,\cdots,x_m)$ 的任意集合 A, B 有

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\Pr\{(X_1,\cdots,X_m)\in A \text{ 和 } (X_{k+1},\cdots,X_{k+m},)\in B\}$$
$$=\Pr\{(X_1,\cdots,X_m)\in A\}\times\Pr\{(X_1,\cdots,X_m)\in B\}.$$

25. 设 $\{\xi_n\}$ 是零均值协方差平稳过程，协方差函数为

$$E[\xi_n\xi_m]=\begin{cases}1, & n=m,\\ \rho, & n\neq m,\end{cases}$$

其中 $0<\rho<1$. 试证明 $\{\xi_n\}$ 有表示式 $\xi_n=U+\eta_n$. 这里 $U, \eta_1, \eta_2, \cdots$ 是零均值不相关的随机变量，$E[U^2]=\rho$ 且 $E[\eta_k^2]=1-\rho$.

提示：利用均方遍历定理于 $U=\lim(\xi_1+\cdots+\xi_n)/n$. 置 $\eta_n=\xi_n-U$ 并计算 $E[U\xi_n], E[U^2]$ 和 $E[\eta_n\eta_m]$.

26. 设 $\{X_n\}$ 是具有有限状态不可约马尔可夫链，其转移概率矩阵为 $\|P_{ij}\|_{i,j=1}^{N}$. 于是，存在一个平稳分布 π，即一个向量 $\pi(1),\cdots,\pi(N)$，满足 $\pi(i)\geqslant 0, i=1,\cdots,N$, $\sum_{i=1}^{N}\pi(i)=1$，且

$$\pi(j)=\sum_{i=1}^{N}\pi(i)P_{ij},\quad j=1,\cdots,N.$$

假设 $\Pr\{X_0=i\}=\pi(i), i=1,\cdots,N$，试证明 $\{X_n\}$ 是弱混合的，从而是遍历的. [533]

27. 设 $\{B(t), t\geqslant 0\}$ 是标准布朗运动，所对应的高斯随机测度为 $B(I)=B(t)-B(s), I=(s,t], 0\leqslant s<t$. 证明若 $f(u), u\geqslant 0$ 是有界连续函数，则

$$Y(t)=\int_0^t f(x)B(\mathrm{d}x),\quad t\geqslant 0$$

是一个鞅.

28. 设 $\{B(t), t\geqslant 0\}$ 是标准布朗运动，所对应的高斯随机测度为 $B(I)=B(t)-B(s), I=(s,t], 0\leqslant s<t$. 设 $f(u)$ 是 $[0,h]$ 上连续函数. 证明

$$Y(t)=\int_t^{t+h} f(u-t)B(\mathrm{d}u),\quad t\geqslant 0$$

是一个平稳过程.

29. 在定理 10.2 条件之下, 证明

$$\lim_{\omega\to\infty} \Pr\{\max_{0\leqslant s\leqslant t/f(\omega)} X(s) < \omega\} = \mathrm{e}^{-t},$$

其中

$$f(w) = \frac{\sqrt{\lambda_2}}{2\pi\sigma}\exp(-\omega^2/2\sigma^2).$$

30. 设 $\{B(t); 0 \leqslant t \leqslant 1\}$ 是标准布朗运动, 所对应的高斯随机测度为 $B(I) = B(t) - B(s), I = (s, t], 0 \leqslant s \leqslant t \leqslant 1$. 试证明恒等式

$$E\left[\exp\left\{\lambda\int_0^1 f(s)\mathrm{d}B(s)\right\}\right] = \exp\left\{\frac{1}{2}\lambda^2\int_0^1 f^2(s)\mathrm{d}s\right\}, \quad -\infty < \lambda < \infty,$$

其中 $f(s), 0 \leqslant s \leqslant 1$, 是连续函数.

31. 设 $\{B(t); 0 \leqslant t \leqslant 1\}$ 是标准布朗运动, 所对应的高斯随机测度为 $B(I) = B(t) - B(s), I = (s, t], 0 \leqslant s \leqslant t \leqslant 1$. 试证明: 若 f 和 g 是满足 $\int_0^1 f(s)g(s)\mathrm{d}s = 0$ 的有界连续函数, 则 $U = \int_0^1 f(s)\mathrm{d}B(s)$ 和 $V = \int_0^1 g(s)\mathrm{d}B(s)$ 是相互独立的随机变量.

附 记

为更好地了解平稳过程的谱理论, 请参阅 Yaglom 的书 [2]. 关于平稳过程的许534 多问题, 包括水平交叉问题, Cramer 和 Leadbetter 已作了专门的讨论 (见 [1]).

参考书目

[1] Cramer, H, and Leadbetter, M. , *Stationary and Related Stochastic Processes*. Wiley, New York, 1966.

[2] Yaglom, A.M., *An Introduction to the Theory of Stationary Random Functions*. Prentice-Hall, Englawood Cliffs, New Jersey, 1962.

535 [3] Anderson, T. W., *Statistical Analysis of Time Series*. Wiley, New York, 1973.

附录 矩阵分析的复习

A.1 谱定理

A. 概念介绍 (线性无关和基)

所有 n 元组 (向量) $\boldsymbol{x}=(x_1,\cdots,x_n)$ 的集合构成一个 n 维向量空间, 其中 x_i 为复数, 定义两个向量 $\boldsymbol{x}=(x_1,\cdots,x_n)$ 和 $\boldsymbol{y}=(y_1,\cdots,y_n)$ 的和 $\boldsymbol{x}+\boldsymbol{y}=(x_1+y_1,\cdots,x_n+y_n)$, 及向量 x 与一个复数 λ 的乘积 $\lambda\boldsymbol{x}=(\lambda x_1,\cdots,\lambda x_n)$.

一个向量组 $\boldsymbol{x}^{(1)},\cdots,\boldsymbol{x}^{(r)}$ 被称为是**线性无关的**, 如果由等式 $c_1\boldsymbol{x}^{(1)}+c_2\boldsymbol{x}^{(2)}+\cdots+c_r\boldsymbol{x}^{(r)}=0$ 可以推出 $c_1=c_2=\cdots=c_r=0$. 例如, 我们给出向量 $(1,0,\cdots,0)$, $(0,1,\cdots,0)$, $\cdots$, 显然它们是线性无关的. 一个 n 元组的线性无关族不可能多于 n 个向量.

设 $\varphi_1,\cdots,\varphi_r(r<n)$ 是一线性无关组, 若存在一个向量 φ_{r+1} 不是 $\varphi_1,\cdots,\varphi_r$ 的线性组合, 即不能表示为 $c_1\varphi_1+\cdots+c_r\varphi_r$ 的形式, 则 $\varphi_1,\cdots,\varphi_r,\varphi_{r+1}$ 线性无关, 继续按这种方式进行下去, 我们可以用向量 $\varphi_{r+1},\cdots,\varphi_n$ 扩充 $\varphi_1,\cdots,\varphi_r$ 使$\varphi_1,\cdots,\varphi_n$ 是一线性无关组, 因为线性无关组至多包含 n 个元素, 所以对每一个向量 $\boldsymbol{y}$ 和每一线性无关组 $\varphi_1,\cdots,\varphi_n$ 都存在常数 $c_1,\cdots,c_n$(且唯一) 使 $\boldsymbol{y}=c_1\varphi_1+\cdots+c_n\varphi_n$.

536

类似的结果也适用于任何线性流形 $\mathfrak{M}$(称向量集 $\mathfrak{M}$ 为线性流形, 若 $\boldsymbol{x},\boldsymbol{y}\in\mathfrak{M}$, 则对任意的复数 a,b 都有 $a\boldsymbol{x}+b\boldsymbol{y}\in\mathfrak{M}$). 对这样一个线性流形 $\mathfrak{M}$, 它的维数是唯一的, 记为整数 $m(0\leqslant m\leqslant n)$, 从而 $\mathfrak{M}$ 中的最大线性无关组包含 m 个元素, 假如 $\varphi_1,\cdots,\varphi_r,r<m$, 是 $\mathfrak{M}$ 中的线性无关组, 则有一向量 $\varphi_{r+1}\in\mathfrak{M}$, 它不能由 $\varphi_1,\cdots,\varphi_r$ 的线性组合来表示, 则我们可以推断存在 $\varphi_{r+1},\cdots,\varphi_m\in\mathfrak{M}$, 使得 $\varphi_1,\cdots,\varphi_m$ 是一线性无关组, 而且对任意的 $\boldsymbol{y}\in\mathfrak{M}$, 存在 (且唯一) 常数 $c_1,\cdots,c_m$, $c_1\varphi_1+\cdots+c_m\varphi_m=\boldsymbol{y}$. 注意到, dim $\mathfrak{M}$=0 表示 $\mathfrak{M}$ 中只包含 0 元素, 而 dim $\mathfrak{M}$=n 表示它包含任何向量, 假如 dim $\mathfrak{M}$=m, $\mathfrak{M}$ 中任何 m 个线性无关向量组都被称为 $\mathfrak{M}$ 的基. 任何 n 个线性无关向量都可以称为 "基".

B. 内积

两个向量 $\boldsymbol{x},\boldsymbol{y}$ 的内积定义为 $(\boldsymbol{x},\boldsymbol{y})=\sum_{i=1}^{n}x_i\bar{y}_i$(其中 $\bar{y}_i$ 为 y_i 的复共轭) , 我们很容易证明下述内积性质:

(i) $(\boldsymbol{x},\boldsymbol{x})\geqslant 0$, 当且仅当 $\boldsymbol{x}=(0,\cdots,0)=0$ 时等式成立.

(ii) $(\lambda \boldsymbol{x}, \boldsymbol{y}) = \lambda(\boldsymbol{x}, \boldsymbol{y})$, 其中 λ 为复数.

(iii) $(\boldsymbol{x}, \boldsymbol{y}) = \overline{(\boldsymbol{y}, \boldsymbol{x})}$, 由 (ii) 得, $(\boldsymbol{x}, \lambda\boldsymbol{y}) = \overline{\lambda}(\boldsymbol{x}, \boldsymbol{y})$.

两个向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 称为正交的, 如果 $(\boldsymbol{x}, \boldsymbol{y})=0$. 一个向量 $\boldsymbol{x}$ 的范数定义为 $\|\boldsymbol{x}\| = (\boldsymbol{x}, \boldsymbol{x})^{\frac{1}{2}}$.

一个复数集 $\{a_{ij}\}(i, j = 1, \cdots, n)$ 定义一个 n 维 (方) 矩阵, 通常表示为 $\boldsymbol{A} = \|a_{ij}\|, i, j = 1, \cdots, n$. 一个 $n \times n$ 矩阵 $\boldsymbol{A}$ 定义 n 维向量空间的一个变换 (或算子), $\boldsymbol{Ax} = \boldsymbol{y}, y_i = \sum_{j=1}^{n} a_{ij}x_j, i = 1, \cdots, n$, 或 $\boldsymbol{xA} = \boldsymbol{z}, z_j = \sum_{i=1}^{n} x_i a_{ij}, j = 1, \cdots, n, j = 1, \cdots, n$. 从这些定义立即可以得到, 对于任意向量 $\boldsymbol{x}, \boldsymbol{y}$ 和常数 α, β:

$$\boldsymbol{A}(\alpha\boldsymbol{x} + \beta\boldsymbol{y}) = \alpha\boldsymbol{Ax} + \beta\boldsymbol{Ay}, \quad (\alpha\boldsymbol{x} + \beta\boldsymbol{y})\boldsymbol{A} = \alpha\boldsymbol{xA} + \beta\boldsymbol{yA},$$

537

进一步, $(\boldsymbol{x}, \boldsymbol{Ay}) = (\boldsymbol{x}\overline{\boldsymbol{A}}, \boldsymbol{y})$, 这里 $\overline{\boldsymbol{A}}$ 是 $n \times n$ 矩阵, 其元素为 $\overline{a}_{ij}$. 这里由右运算和左运算产生的两个变换是相互对偶的. 我们可以在大部分矩阵论的教科书上看到更加详细的线性变换的几何理论和代数理论.

C. 特征值和特征向量

一个复数 λ 被称为矩阵 $\boldsymbol{A}$ 的特征值, 如果存在一个向量 $\boldsymbol{x}^{(\lambda)} \neq \boldsymbol{0}$ 使得 $\boldsymbol{Ax}^{(r)} = \lambda\boldsymbol{x}^{(r)}$. 如果 λ 是 $\boldsymbol{A}$ 的一个特征值, 我们把满足方程 $\boldsymbol{Ax} = \lambda\boldsymbol{x}$ 的所有向量集记为 $\mathfrak{M}_\lambda$, 且把它称为 $\boldsymbol{A}$ 对应于特征值 λ 的右特征流形, 把 $\mathfrak{M}_\lambda$ 中的向量称为对应于 λ 的右特征向量. 显然, 对任意的 $\boldsymbol{y}, \boldsymbol{z} \in \mathfrak{M}_\lambda$, 有 $a\boldsymbol{y} + b\boldsymbol{z} \in \mathfrak{M}_\lambda$, 其中 a, b 为任意常数. $\mathfrak{M}_\lambda$ 的维数被称为 λ 的几何重数.

如果 $\mathfrak{M}_1, \mathfrak{M}_2, \cdots, \mathfrak{M}_r$ 是算子 $\boldsymbol{A}$ 互不相同的特征流形, 且 $\varphi_1, \cdots, \varphi_r$ 是分别属于 $\mathfrak{M}_1, \mathfrak{M}_2, \cdots, \mathfrak{M}_r$ 的任意非零向量, 那么 $\varphi_1, \cdots, \varphi_r$ 是线性无关的. 事实上, 反之, 假设 m 是我们能够找到的最小整数, 使得 $\mathfrak{M}_1, \mathfrak{M}_2, \cdots, \mathfrak{M}_m$ 互异, 则存在非零向量 $\varphi_1 \in \mathfrak{M}_1, \varphi_2 \in \mathfrak{M}_2, \cdots, \varphi_m \in \mathfrak{M}_m$, 和非零常数 $c_1, \cdots, c_m$ 使得 $c_1\varphi_1 + \cdots + c_m\varphi_m = \boldsymbol{0}$. 一般地, $m \geqslant 2$. 算子 $\boldsymbol{A}$ 作用到上式两边, 可以得到 $\lambda_1 c_1\varphi_1 + \cdots + \lambda_m c_m\varphi_m = \boldsymbol{0}$, 其中 λ_i 为对应于特征流形 $\mathfrak{M}_i$ 的特征值. 如果 λ_i 中有一个为零, 我们得到 $(m-1)$ 个线性无关元素, 这与 m 的定义相矛盾. 因此, $\lambda_1 \neq 0$; 用 λ_1 乘以 $c_1\varphi_1 + \cdots + c_m\varphi_m = \boldsymbol{0}$ 之后再与 $\lambda_1 c_1\varphi_1 + \cdots + \lambda_m c_m\varphi_m = \boldsymbol{0}$ 相减, 我们有

$$(\lambda_2 - \lambda_1)c_2\varphi_2 + \cdots + (\lambda_m - \lambda_1)c_m\varphi_m = \boldsymbol{0}.$$

这又与定义的最小 m 相矛盾. 假设 $\mathfrak{M}_1, \mathfrak{M}_2, \cdots, \mathfrak{M}_r$ 是 $\boldsymbol{A}$ 的互为相异的特征流形, 而且 $\varphi_1^{(i)}, \cdots, \varphi_{m_i}^{(i)}$ 是 $\mathfrak{M}_i$ 的一组基, $i = 1, \cdots, r$, 那么

$$\varphi_1^{(1)}, \cdots, \varphi_{m_1}^{(1)}, \quad \varphi_1^{(2)}, \cdots, \varphi_{m_2}^{(2)}, \cdots, \quad \varphi_1^{(r)}, \cdots, \varphi_{m_r}^{(r)}$$

构成一个线性无关组. 因此, $\boldsymbol{A}$ 只有有限个特征值和特征流形. 在特征流形维数的和为 n 的条件下, 我们可以得到完全由特征向量构成的一组基 (对于全空间), 具有这种特性的矩阵被称为可对角化矩阵.

我们同样可以用方程 $\boldsymbol{xA} = \lambda\boldsymbol{x}$ 替代 $\boldsymbol{Ax} = \lambda\boldsymbol{x}$. 同理, 满足 $\boldsymbol{xA} = \lambda\boldsymbol{x}$ 的 λ 值有一个非平凡解, 它正好是 $\boldsymbol{A}$ 的特征值 (已在前面段落定义). 进一步, 满足 $\boldsymbol{xA} = \lambda\boldsymbol{x}$ 的所有向量 (即左特征向量) 形成的流形的维数也正是 λ 的重数 (读者可以证明这一结论). 和前面一样, 由不同特征值的左特征向量构成的集合是线性无关的. 538

众所周知, $\boldsymbol{A}$ 的特征值正好是 n 次代数方程 $\det\|\boldsymbol{A} - \lambda\boldsymbol{I}\| = 0$ 的根. 因此, 根据行列式性质, 有一个结果下面将要用到: 如果 $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_1 & \boldsymbol{0} \\ \boldsymbol{B} & \boldsymbol{A}_2 \end{pmatrix}$, 其中 $\boldsymbol{A}_1, \boldsymbol{A}_2$ 是方阵, 则 λ 是$\boldsymbol{A}$的一个特征值当且仅当 λ 是矩阵 $\boldsymbol{A}_1$ 或 $\boldsymbol{A}_2$ 的一个特征值. 事实上, 因为

$$\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = \det(\boldsymbol{A}_1 - \lambda\boldsymbol{I})\det(\boldsymbol{A}_2 - \lambda\boldsymbol{I}),$$

其中同样的记号 $\boldsymbol{I}$ 用来表示不同维数的单位矩阵, 这个结论是显然的.

(a) 谱表示

在下面讨论中, 我们假设 $\boldsymbol{A}$ 是实矩阵, 即 $\boldsymbol{A}$ 的所有元素都是实数. 假设我们可以用矩阵 $\boldsymbol{A}$ 的右特征向量构造一组基. 由前面的讨论可知, 我们也可以用矩阵 $\boldsymbol{A}$ 的左特征向量构造一组基. 另外, 如果 $\boldsymbol{A}$ 的元素 a_{ij} 都是实的, 我们可以选择两组正交基, 即 $\boldsymbol{\varphi}_1, \cdots, \boldsymbol{\varphi}_n$ 和 $\boldsymbol{\psi}_1, \cdots, \boldsymbol{\psi}_n$ 分别是由右特征向量和左特征向量构成的两组基, 而且当 $i = j$ 时 $(\boldsymbol{\varphi}_i, \boldsymbol{\psi}_j) = 1$, 当 $i \neq j$ 时 $(\boldsymbol{\varphi}_i, \boldsymbol{\psi}_j) = 0$. 为了证明具有这些性质的特征向量的构造, 我们首先注意到, 如果 $\boldsymbol{Ax}_i = \lambda_i\boldsymbol{x}_i$ 和 $\boldsymbol{y}_j\boldsymbol{A} = \mu_j\boldsymbol{y}_j$ 成立, 则

$$\mu_j(\boldsymbol{y}_j, \boldsymbol{x}_i) = (\mu_j\boldsymbol{y}_j, \boldsymbol{x}_i) = (\boldsymbol{y}_j\boldsymbol{A}, \boldsymbol{x}_i) = (\boldsymbol{y}_j, \boldsymbol{A}x_i) = (\boldsymbol{y}_j, \lambda_i\boldsymbol{x}_i) = \overline{\lambda}_i(\boldsymbol{y}_j, \boldsymbol{x}_i),$$

因此, 假如 $\mu_j \neq \overline{\lambda}_i$, 则一定有 $(\boldsymbol{y}_j, \boldsymbol{x}_i)$=0. 注意到, 由于 $\boldsymbol{A}$ 是实的, 直接可以推得, 当 $\boldsymbol{A}x = \lambda\boldsymbol{x}$ 成立时, $\boldsymbol{A}\overline{\boldsymbol{x}} = \overline{\lambda}\overline{\boldsymbol{x}}$ 也成立, 其中 $\overline{\boldsymbol{x}} = (\overline{x}_1, \cdots, \overline{x}_n)$. 所以, 我们得到 $\boldsymbol{A}$ 的特征值互为共轭且成对出现, 并且 λ 与 $\bar{\lambda}$ 有相同的重数. 设 $\boldsymbol{A}$ 的特征值是

$$\lambda_1, \bar{\lambda}_1, \lambda_2, \bar{\lambda}_2, \cdots, \lambda_r, \bar{\lambda}_r, \lambda_{r+1}, \lambda_{r+2}, \cdots \lambda_m,$$

其中 $\lambda_1, \cdots, \lambda_r$ 是复数, $\lambda_{r+1}, \cdots, \lambda_m$ 是实数. 我们分别用 $\mathfrak{M}_1, \overline{\mathfrak{M}}_1, \cdots, \mathfrak{M}_r, \overline{\mathfrak{M}}_r, \mathfrak{M}_{r+1}, \cdots, \mathfrak{M}_m$ 表示相应的右特征流形, 用 $\mathfrak{L}_1, \overline{\mathfrak{L}}_1, \cdots, \mathfrak{L}_r, \overline{\mathfrak{L}}_r, \mathfrak{L}_{r+1}, \cdots, \mathfrak{L}_m$ 表示相应的左特征流形. 539

现在, 我们已经证明 $\mathfrak{L}_1$ 中的每个向量正交于右特征流形中除了 $\overline{\mathfrak{M}}_1$ 的每个向量, 并且 (类似地) 正交于其他的左特征流形中的每个向量. 因此, 我们的任务是选择 $\mathfrak{L}_1$ 中的一组基 $\boldsymbol{\psi}_1, \cdots, \boldsymbol{\psi}_d$ 以及 $\overline{\mathfrak{M}}_1$ 中的一组基 $\boldsymbol{\varphi}_1, \cdots, \boldsymbol{\varphi}_d$, 使得当 $i = j$ 时

$(\psi_i,\varphi_j)=1$, 当 $i\neq j$ 时 $(\psi_i,\varphi_j)=0$, 其中 d 为 λ_1 的重数, 类似地, 对 $\overline{\mathfrak{L}}_1$, $\overline{\mathfrak{M}}_1$, $\mathfrak{L}_2$, $\overline{\mathfrak{M}}_2$ 做相同的处理. 为此, 任取 $\overline{\mathfrak{M}}_1$ 中的一组基 $\varphi_1,\cdots,\varphi_d$, 以及 $\overline{\mathfrak{L}}_1$ 中的一组基 $\boldsymbol{y}_1,\cdots,\boldsymbol{y}_d$. 我们希望能找到一组常数 $c_1,\cdots,c_d$ 使得 $\psi_1=c_1\boldsymbol{y}_1+\cdots+c_d\boldsymbol{y}_d$, 并且满足条件 $(\psi_1,\varphi_1)=1$ 和 $(\psi_1,\varphi_i)=0$, $i=2,\cdots,d$; 因此, 我们希望它们满足如下方程组

$$\begin{aligned}
c_1(\boldsymbol{y}_1,\varphi_1)+c_2(\boldsymbol{y}_2,\varphi_1)+\cdots+c_d(\boldsymbol{y}_d,\varphi_1)&=1,\\
c_1(\boldsymbol{y}_1,\varphi_2)+c_2(\boldsymbol{y}_2,\varphi_2)+\cdots+c_d(\boldsymbol{y}_d,\varphi_2)&=0,\\
\vdots\qquad\qquad\vdots\qquad\qquad\qquad\vdots\quad&\\
c_1(\boldsymbol{y}_1,\varphi_d)+c_2(\boldsymbol{y}_2,\varphi_d)+\cdots+c_d(\boldsymbol{y}_d,\varphi_d)&=0.
\end{aligned}$$

假设这个关于 $c_1,\cdots,c_d$ 的方程组无解. 这意味着 $(1,0,\cdots,0)$ 无法表示成 d 个向量 $\boldsymbol{f}_1=((\boldsymbol{y}_1,\varphi_1),\cdots,(\boldsymbol{y}_1,\varphi_d)),\cdots,f_d=((\boldsymbol{y}_d,\varphi_1),\cdots,(\boldsymbol{y}_d,\varphi_d))$ 的线性组合. 所以, $\boldsymbol{f}_1,\cdots,\boldsymbol{f}_d$ 是线性相关的. 因此, 存在一组不全为零的常数 $a_1,\cdots,a_d$ 使得 $a_1\boldsymbol{f}_1+\cdots+a_d\boldsymbol{f}_d=0$. 这导致方程

$$(a_1\boldsymbol{y}_1+\cdots+a_d\boldsymbol{y}_d,\varphi_i)=0,i=1,\cdots d.$$

成立. 但是, 这证明了向量组 $\boldsymbol{y}_1,\cdots,\boldsymbol{y}_d$(包括 y_i 的任意线性组合) 正交于右特征流形中 (除了 $\overline{\mathfrak{M}}_1$) 的任一向量. 现在, 已知 $a_1\boldsymbol{y}_1+\cdots+a_d\boldsymbol{y}_d$ 正交于任意右特征向量以及任意右特征向量的线性组合. 但是, 根据假设, 存在由右特征向量构成的一组基, 使得 $a_1\boldsymbol{y}_1+\cdots+a_d\boldsymbol{y}_d$ 正交与它本身, 因此 $a_1\boldsymbol{y}_1+\cdots+a_d\boldsymbol{y}_d=0$. 这与 $\boldsymbol{y}_1,\cdots,\boldsymbol{y}_d$ 线性无关相矛盾. 因此, 存在满足条件的 ψ_1. 我们用同样的方法构造 $\psi_2,\cdots,\psi_d$. 我们仍须证明 $\psi_1,\cdots,\psi_d$ 是线性无关的. 假设 $a_1\psi_1+\cdots+a_d\psi_d=0$, 则

$$\begin{aligned}
0&=(a_1\psi_1+\cdots+a_d\psi_d,\varphi_1)=a_1\\
0&=(a_1\psi_1+\cdots+a_d\psi_d,\varphi_2)=a_2\\
&\ \ \vdots\\
0&=(a_1\psi_1+\cdots+a_d\psi_d,\varphi_d)=a_d
\end{aligned}$$

540 所以, $\psi_1,\cdots,\psi_d$ 是线性无关的.

众所周知, 如果 $\boldsymbol{A}$ 是一个实矩阵, 并且它的右 (或左) 特征向量可以作为全空间的一组基, 我们可以取右特征向量构成的一组基 $\varphi_1,\cdots,\varphi_n$ 以及左特征向量构成的一组基 $\psi_1,\cdots,\psi_n$, 使得它们是相互双正交的, 即满足当 $i=j$ 时, $(\psi_i,\varphi_j)=1$; 当 $i\neq j$ 时, $(\psi_i,\varphi_j)=0$.

我们可以运用这个结论展开矩阵 $\boldsymbol{A}$ 的典型表示, 即所谓矩阵的谱表示. 假设 $\lambda_1,\cdots,\lambda_n$ 是对应于特征向量 $\varphi_1,\cdots,\varphi_n$ 的右特征值, 即分别使得 $\boldsymbol{A}\varphi_i=\lambda_i\varphi_i$,

$i=1,\cdots,n$, 其中 λ_i 不必互异. 令

$$\boldsymbol{\varphi}_i=(\varphi_{i1},\cdots,\varphi_{in}),\qquad \boldsymbol{\psi}_i=(\psi_{i1},\cdots,\psi_{in}),$$

$$\boldsymbol{\Phi}=\begin{pmatrix}\varphi_{11} & \cdots & \varphi_{n1}\\ \vdots & \ddots & \vdots\\ \varphi_{1n} & \cdots & \varphi_{nn}\end{pmatrix},\quad \boldsymbol{\Psi}=\begin{pmatrix}\psi_{11} & \cdots & \psi_{1n}\\ \vdots & \ddots & \vdots\\ \psi_{n1} & \cdots & \psi_{nn}\end{pmatrix},$$

$$\boldsymbol{\Lambda}=\begin{pmatrix}\lambda_1 & 0 & \cdots & \mathbf{0}\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & 0 & \cdots & \lambda_n\end{pmatrix}.$$

因为 $\varphi_1,\cdots,\varphi_n$ 和 $\psi_1,\cdots,\psi_n$ 是双正交的, 所以 $\boldsymbol{\Psi\Phi}=\boldsymbol{I}$, 其中 $\boldsymbol{I}$ 是单位矩阵. 进一步, 我们可以直接计算得 $\boldsymbol{\Phi\Lambda\Psi}\boldsymbol{\varphi}_i=\lambda_i\boldsymbol{\varphi}_i$, $i=1,\cdots,n$. 因为 $\boldsymbol{A}\boldsymbol{\varphi}_i=\lambda_i\boldsymbol{\varphi}_i$, $i=1,\cdots,n$, 并且 φ_i 是全空间的一组基, 故有

$$\boldsymbol{A}=\boldsymbol{\Phi\Lambda\Psi}\ 且\ \boldsymbol{\Phi\Psi}=\boldsymbol{I}.$$

由此可得, $\boldsymbol{A}^m=\boldsymbol{\Phi\Lambda\Psi\Phi\Lambda\Psi}\cdots\boldsymbol{\Phi\Lambda\Psi}=\boldsymbol{\Phi\Lambda}^m\boldsymbol{\Psi}$, 其中

$$\boldsymbol{\Lambda}^m=\begin{pmatrix}\lambda_1^m & 0 & \cdots & \mathbf{0}\\ 0 & \lambda_2^m & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & 0 & \cdots & \lambda_n^m\end{pmatrix},$$

所以, 如果矩阵 $\boldsymbol{A}$ 的谱表示是已知的, 则 $\boldsymbol{A}^m$ 是相对容易计算的.

(b) 收敛

我们需要建立向量序列和矩阵序列收敛的概念. [541]

给定 n 维空间中的一个向量序列 $\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)},\cdots$, 称该序列收敛于 $\boldsymbol{x}^{(0)}$, 如果

$$\lim_{j\to\infty}x_i^{(j)}=x_i^{(0)}, i=1,\cdots,n.$$

类似地, 给定一个 n 维方阵序列 $\boldsymbol{A}^{(1)},\boldsymbol{A}^{(2)},\cdots$, 称该序列收敛于矩阵 $\boldsymbol{A}^{(0)}$, 如果

$$\lim_{h\to\infty}a_{ij}^{(h)}=a_{ij}^{(0)}, i,j=1,\cdots,n.$$

由此定义, 我们可以得到, 如果 $\lim\limits_{h\to\infty}\boldsymbol{A}^{(h)}=\boldsymbol{A}^{(0)}$ 以及 $\lim\limits_{j\to\infty}\boldsymbol{x}^{(j)}=\boldsymbol{x}^{(0)}$, 那么 $\lim\limits_{h\to\infty}\boldsymbol{A}^{(h)}\boldsymbol{x}^{(h)}=\boldsymbol{A}^{(0)}\boldsymbol{x}^{(0)}$. 进一步, 如果存在矩阵 $\boldsymbol{A}^{(0)}$ 以及全空间的一组基 $\boldsymbol{x}^{(1)}$,

$\boldsymbol{x}^{(2)},\cdots,\boldsymbol{x}^{(n)}$, 使得对于矩阵序列 $\{\boldsymbol{A}^{(j)}\}$,

$$\lim_{h\to\infty}\boldsymbol{A}^{(h)}\boldsymbol{x}^{(i)}=\boldsymbol{A}^{(0)}\boldsymbol{x}^{(i)}, i=1,\cdots,n.$$

那么, 就有 $\lim\limits_{h\to\infty}\boldsymbol{A}^{(h)}=\boldsymbol{A}^{(0)}$. 事实上, 对任意 $\boldsymbol{y}=c_1\boldsymbol{x}^{(1)}+\cdots+c_n\boldsymbol{x}^{(n)}$, 都有 $\lim\limits_{h\to\infty}\boldsymbol{A}^{(h)}\boldsymbol{y}=\boldsymbol{A}^{(0)}\boldsymbol{y}$, 因为 $\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)},\cdots\boldsymbol{x}^{(n)}$ 是全空间的一组基.

D. 正定矩阵

一个 $n\times n$ 阶的实矩阵 $\boldsymbol{A}=(a_{ij})$ 是正定的, 如果 $\sum\limits_{i,j}a_{ij}x_ix_j>0$, 除非每个 $x_i=0$. 在大多数情况, 我们只考虑对称的正定阵, 即 $a_{ij}=a_{ji}$.

一个实对称的正定阵是非奇异的, 并且其所有特征值都是正的. 每个这种矩阵都有一个 "平方根", 即存在一个非奇异的实矩阵 $\boldsymbol{P}=(p_{ij})$, 使得 $\boldsymbol{A}=\boldsymbol{P}\boldsymbol{P}^{\mathrm{T}}$. 这里, $\boldsymbol{P}^{\mathrm{T}}$ 表示矩阵 $\boldsymbol{P}$ 的转置, 其元素 $p_{ij}^{\mathrm{T}}=p_{ji}$.

A.2 正定矩阵的 Frobenius 理论

正矩阵的 Frobenius 理论大量地应用于概率论中, 特别是在马尔可夫转移矩阵的分析中. 我们给出该理论在下面几个方面的应用.

预备知识

542 假设 $\boldsymbol{A}=(a_{ij})(i,j=1,\cdots,n)$ 是一个方阵. 如果每个 a_{ij} 都是非负的, 我们记 $\boldsymbol{A}\geqslant\boldsymbol{0}$; 如果 $\boldsymbol{A}\geqslant\boldsymbol{0}$, 且至少存在一个 $a_{ij}>0$, 则我们记 $\boldsymbol{A}>\boldsymbol{0}$, 并且称 $\boldsymbol{A}$ 是正矩阵; 如果每个 a_{ij} 都是正的, 我们记 $\boldsymbol{A}\gg\boldsymbol{0}$. 对于向量 $\boldsymbol{x}=(x_1,\cdots,x_n)$, 我们引入相同的记号, 即当 $x_i\geqslant 0(i=1,\cdots,n)$ 时, 记 $\boldsymbol{x}\geqslant\boldsymbol{0}$; 当 $\boldsymbol{x}\geqslant\boldsymbol{0}$ 并且至少存在一个 $x_i>0$ 时, 记 $\boldsymbol{x}>\boldsymbol{0}$; 当 $x_i>0(i=1,\cdots,n)$ 时, 记 $\boldsymbol{x}\gg\boldsymbol{0}$. 同样地, 当 $\boldsymbol{x}-\boldsymbol{y}\geqslant\boldsymbol{0}$ 时, 记 $\boldsymbol{x}\geqslant\boldsymbol{y}$, 等等. 显然, 若 $\boldsymbol{A}\geqslant\boldsymbol{0}$ 且 $\boldsymbol{x}\geqslant\boldsymbol{y}$, 则 $\boldsymbol{A}\boldsymbol{x}\geqslant\boldsymbol{A}\boldsymbol{y}$, 若 $\boldsymbol{A}\gg\boldsymbol{0}$ 且 $\boldsymbol{x}>\boldsymbol{y}$, 有 $\boldsymbol{A}\boldsymbol{x}\gg\boldsymbol{A}\boldsymbol{y}$.

令 $\boldsymbol{A}\geqslant\boldsymbol{0}$, 且 Λ 是由所有对应于向量 $\boldsymbol{x}=(x_1,\cdots,x_n)$ 的实数 λ 构成的集合, 使得

$$\sum_{i=1}^{n}x_i=1,\quad \boldsymbol{x}>0,\text{ 且 }\quad \boldsymbol{A}\boldsymbol{x}\geqslant\lambda\boldsymbol{x}.$$

令 $\lambda_0=\sup_\Lambda\lambda$; 则 λ_0 是有限的, 而且容易证明, 如果 $\boldsymbol{A}\gg\boldsymbol{0}$, 则 λ_0 是正的. 事实上, 如果 $M=\max_{1\leqslant i,j\leqslant n}a_{ij}$ 则由 $\sum\limits_{i=1}^{n}x_i=1$ 且 $\boldsymbol{x}>\boldsymbol{0}$ 可推得 $\sum\limits_{j=1}^{n}a_{ij}x_j\leqslant M\sum\limits_{j=1}^{n}x_j=M$, $i=1,\cdots,n$, 而且至少存在一个 $x_j\geqslant\dfrac{1}{n}$. 则有 $\lambda_0\leqslant n\mathrm{M}$. 类似地, 如果 $\boldsymbol{A}\gg\boldsymbol{0}$ 且

$\boldsymbol{x}>\mathbf{0}$, 则 $0<\delta=\min_{1\leqslant i,j\leqslant n}a_{ij}$, 且 $\sum_{j=1}^{n}a_{ij}\left(\frac{1}{n}\right)\geqslant\delta$, $i=1,\cdots,n$, 从而 $\lambda_0\geqslant\delta n$.

假设对于一个矩阵 $\boldsymbol{A}>\mathbf{0}$, 我们有 $\lambda_0=\mathbf{0}$. 如果 $\boldsymbol{x}\gg\mathbf{0}$, 由 $\lambda_0=0$ 可知, $\boldsymbol{Ax}\gg\mathbf{0}$ 是不成立的. 因为对某个 $\boldsymbol{x}\gg\mathbf{0}$, $\boldsymbol{Ax}=\mathbf{0}$, 则要求 $\boldsymbol{A}=\mathbf{0}$. 于是, 存在某个 $\boldsymbol{x}\gg\mathbf{0}$, 使得 $\boldsymbol{Ax}>\mathbf{0}$. 令 C_1 为 $\boldsymbol{Ax}$ 正元素的指标集, 显然 C_1 不依赖于 $\boldsymbol{x}\gg\mathbf{0}$ 的选择. 令 $\boldsymbol{y}=(y_1,\cdots,y_n)>0$, 使得若 $i\in C_1$, 则 $y_i>0$, 若 $i\notin C_1$, 则 $y_i=0$, 且定义 C_2 为 $\boldsymbol{Ay}$ 正元素的指标集. 同样地, C_2 也不依赖于 $\boldsymbol{y}$ 的选取, 且 $C_2\subseteq C_1$. 因为 $\lambda_0=0$, 我们可得 $C_2\neq C_1$. 按照这样的方法继续下去, 我们可以找到

$$C_1\supset C_2\supset\cdots\supset C_m=C_{m+1}=\cdots=\theta,$$

这个结论是恰当的. 现在, 我们可以得到 $\boldsymbol{A}^m=\mathbf{0}$. 事实上, 显然对任意的 $\boldsymbol{x}\gg\mathbf{0}$, 都有 $\boldsymbol{A}^m\boldsymbol{x}=\mathbf{0}$, 因为每个向量都可以写成两个严格正向量的差, 对任意 $\boldsymbol{z}$, 都有 $\boldsymbol{A}^m\boldsymbol{z}=\mathbf{0}$, 这等价于 $\boldsymbol{A}^m=\mathbf{0}$.

Frobenius 第一定理 现在, 我们将证明 Frobenius 第一定理.

定理 2.1 如果 $\boldsymbol{A}\gg\mathbf{0}$, 则

(a) 存在 $\boldsymbol{x}^0\gg\mathbf{0}$, 使得 $\boldsymbol{Ax}^0=\lambda_0\boldsymbol{x}^0$;

(b) 如果 $\lambda\neq\lambda_0$ 是 $\boldsymbol{A}$ 的任一其他特征值, 则 $|\lambda|<\lambda_0$;

(c) $\boldsymbol{A}$ 的特征值 λ_0 的右特征向量形成一个一维子空间, 即 $\dim\mathfrak{M}_{\lambda_0}=1$. [543]

证明 (a) 由 λ_0 的定义, 存在一列 $\gamma_1,\gamma_2,\cdots\to\lambda_0$, 且存在一列向量 $x^{(1)},x^{(2)},\cdots$, 使得

$$\boldsymbol{x}^{(i)}>0, \boldsymbol{Ax}^{(i)}\geqslant\gamma_i\boldsymbol{x}^{(i)},\quad x_1^{(i)}+\cdots+x_n^{(i)}=1. \tag{A.2.1}$$

由于所有 $\boldsymbol{x}^{(i)}$ 的坐标都落在区间 [0, 1] 内, 我们可以通过对角化得到一列正整数 $n_1<n_2<n_3<\cdots$, 和一个向量 $\boldsymbol{x}^{(0)}=(x_1^{(0)},\cdots,x_n^{(0)})$, 其中 $x_r^{(0)}\in[0,1]$, $(r=1,2,\cdots)$, 使得

$$\lim_{j\to\infty}x_r^{(n_j)}\to x_r^0,\quad r=1,\cdots,n. \tag{A.2.2}$$

由 (A.2.1) 可以得到 $x_1^0+\cdots+x_n^0=1$, 且 $\boldsymbol{x}^0>\mathbf{0}$. 如果用 n_j 代替 (A.2.1) 第二个不等式中的 i, 令 j→ ∞, 则 $\boldsymbol{Ax}^0\geqslant\lambda_0\boldsymbol{x}^0$. 因此, 有 $\boldsymbol{Ax}^0=\lambda_0\boldsymbol{x}^0$, 否则 $\boldsymbol{Ax}^0>\lambda_0\boldsymbol{x}^0$. 将 $\boldsymbol{A}$ 作用于上个不等式的两边, 注意到 $\boldsymbol{A}\gg0$, 令 $\boldsymbol{y}^0=\boldsymbol{Ax}^0$, 则有 $\boldsymbol{Ay}^0\gg\lambda_0\boldsymbol{y}^0$ 且 $\boldsymbol{y}^0\gg0$. 因此, 对于充分小的 $\varepsilon>0$, 则有 $\boldsymbol{Ay}^0\gg(\lambda_0+\varepsilon)\boldsymbol{y}^0$, 用适当的正数乘以 $\boldsymbol{y}^0$, 使得它的各个坐标的和都等于 1, 所以 $(\lambda_0+\varepsilon)\in\Lambda$, 这与 λ_0 的定义相矛盾. 因此, $\boldsymbol{Ax}^0=\lambda_0\boldsymbol{x}^0$. 由于 $\boldsymbol{x}^0>\mathbf{0}$ 且 $\boldsymbol{A}\gg\mathbf{0}$, 我们得到 $\lambda_0\boldsymbol{x}^0\gg\mathbf{0}$, 或者 $\boldsymbol{x}^0\gg\mathbf{0}$, 这就证明了 (a).

(b) 假设 $\lambda \neq \lambda_0$ 且 $\boldsymbol{Az} = \lambda \boldsymbol{z}$, 其中 $\boldsymbol{z} \neq \boldsymbol{0}$. $\boldsymbol{Az} = \lambda \boldsymbol{z}$ 的坐标形式为

$$\sum_{j=1}^{n} a_{ij} z_j = \lambda z_i, \quad i = 1, \cdots, n.$$

两边同时取绝对值, 因为 $a_{ij} > 0$, 再利用和的绝对值不会超过绝对值的和, 则有

$$\sum_{j=1}^{n} a_{ij} |z_j| \geqslant |\lambda||z_i|, \quad i = 1, \cdots, n,$$

即

$$\boldsymbol{A}|\boldsymbol{z}| \geqslant |\lambda||\boldsymbol{z}|, \quad \text{其中 } |z| = (|z_1|, \cdots, |z_n|).$$

用适当的正常数乘以 $|\boldsymbol{z}|$, 使得它的各个坐标的和等于 1(又 $\boldsymbol{z} \neq \boldsymbol{0}$), 所以 $|\lambda| \in \Lambda$, 因此, 由 λ_0 的定义可知, $|\lambda| \leqslant \lambda_0$. 为了证明 $|\lambda| < \lambda_0$, 考虑矩阵 $\boldsymbol{A}_\delta = \boldsymbol{A} - \delta \boldsymbol{I}$, 这里的 $\boldsymbol{I}$ 是单位矩阵, 取 δ 充分小, 使得 $\boldsymbol{A}_\delta \gg \boldsymbol{0}$. 因为 λ_0 是 $\boldsymbol{A}$ 的最大正特征值, 则 $\lambda_0 - \delta$ 是 $\boldsymbol{A}_\delta$ 的最大正特征值.

544

类似地, 对于 $|\lambda| \leqslant \lambda_0$, 用 $\boldsymbol{A}_\delta$ 和 $\lambda - \delta$ 分别代替上述结论中的 $\boldsymbol{A}$ 和 λ, 得到 $|\lambda - \delta| \leqslant \lambda_0 - \delta$. 但是

$$|\lambda| = |\lambda - \delta + \delta| \leqslant |\lambda - \delta| + \delta \leqslant \lambda_0,$$

所以, $|\lambda| = \lambda_0$ 意味着 $|\lambda| = |\lambda - \delta| + \delta$, 这就要求 λ 为正实数. 因此, $\lambda = |\lambda| = \lambda_0$, 这与假设 $\lambda \neq \lambda_0$ 相矛盾.

(c) 假设 $\boldsymbol{Ay} = \lambda_0 \boldsymbol{y}$ 且不存在常数 c 使得 $\boldsymbol{y} = c\boldsymbol{x}^0$. 由于 $\boldsymbol{A}$ 是实矩阵, 向量 $\boldsymbol{u}, \boldsymbol{v}$ 的坐标分别由实数和虚数部分组成, $\boldsymbol{y}$ 的坐标是 $\boldsymbol{A}$ 的特征值 λ_0 所对应的特征向量, 且对于任意的 c 都有 $\boldsymbol{y} \neq c\boldsymbol{x}^0$, 则 $\boldsymbol{u}, \boldsymbol{v}$ 至少有一个不是 $c\boldsymbol{x}^0$ 的形式. 这样, 我们首先不妨设 $\boldsymbol{y}$ 是实的. 因为 $\boldsymbol{x}^0 \gg \boldsymbol{0}$, 取适当的 μ, 如令 $|\mu| = \min_{y_i \neq 0} \{x_i^0 / |y_i|\}$, 使得 $\boldsymbol{x}^0 - \mu \boldsymbol{y} > \boldsymbol{0}$, 但不是 $\gg \boldsymbol{0}$; 但是, $\boldsymbol{A}(\boldsymbol{x}^0 - \mu \boldsymbol{y}) = \lambda_0 (\boldsymbol{x}^0 - \mu \boldsymbol{y})$, 正如 (a) 的证明, 必有 $(\boldsymbol{x}^0 - \mu \boldsymbol{y}) \gg \boldsymbol{0}$, 这与 μ 的选择相矛盾. ■

我们再来看一些简单的结论. 如果 $\boldsymbol{A} \gg \boldsymbol{0}$, 则存在向量 $\boldsymbol{f}^0 \gg \boldsymbol{0}$ 使得 $\boldsymbol{f}^0 \boldsymbol{A} = \lambda_0 \boldsymbol{f}^0$, 且与 λ_0 相对应的左特征向量的流形是一维的. 为此, 令 $\lambda' = \sup_{\Lambda'} \lambda$, 其中 $\Lambda' = \{\lambda | \boldsymbol{fA} \geqslant \lambda \boldsymbol{f}, \boldsymbol{f} > \boldsymbol{0}\}$, 正如定理 2.1 的证明, 可知存在 $\boldsymbol{f}^0 \gg \boldsymbol{0}$ 使得 $\boldsymbol{f}^0 \boldsymbol{A} = \lambda' \boldsymbol{f}^0$, 若 λ 是不等于 λ' 的特征值, 则 $|\lambda| < \lambda'$, 且与 λ' 相对应的左特征向量的流形是一维的. 这就意味着, 若 $\lambda_0 \neq \lambda'$, 则 $|\lambda_0| < \lambda'$, 因为 λ_0 是 $\boldsymbol{A}$ 的特征值. 但是, 定理 2.1 指出, 如果特征值 $\lambda' \neq \lambda_0$, 则 $|\lambda'| < \lambda_0$. 因此, $\lambda' = \lambda_0$.

定理 2.2　如果 $\boldsymbol{A} > \boldsymbol{0}$, 且对于某些整数 $m > 0$, $\boldsymbol{A}^m \gg \boldsymbol{0}$, 则上述定理仍然成立.

证明 如定理 2.1 的证明, 我们可以找到 $\boldsymbol{x}^0 > \mathbf{0}$ 使得 $\boldsymbol{A}\boldsymbol{x}^0 \geqslant \lambda_0\boldsymbol{x}^0$. 如果 $\boldsymbol{A}\boldsymbol{x}^0 \neq \lambda_0\boldsymbol{x}^0$, 则 $\boldsymbol{A}\boldsymbol{x}^0 > \lambda_0\boldsymbol{x}^0$. 将 $\boldsymbol{A}^m$ 作用于两边, 得到 $\boldsymbol{A}^{m+1}\boldsymbol{x}^0 \gg \lambda_0\boldsymbol{A}^m\boldsymbol{x}^0$ 和 $\boldsymbol{y} = \boldsymbol{A}^m\boldsymbol{x}^0 \gg \mathbf{0}$. 因此, $\boldsymbol{A}\boldsymbol{y} \gg \lambda_0\boldsymbol{y}$, 根据定理 2.1 的证明, 这与 λ_0 的定义相矛盾; 因此, $\boldsymbol{A}\boldsymbol{x}^0 = \lambda_0\boldsymbol{x}^0$. 将 $\boldsymbol{A}$ 连续作用于 $\boldsymbol{A}\boldsymbol{x}^0 = \lambda_0\boldsymbol{x}^0$ 的两边, 得到 $\boldsymbol{A}^m\boldsymbol{x}^0 = \lambda_0^m\boldsymbol{x}^0$. 因为 $\boldsymbol{A}^m \gg \mathbf{0}$, $\boldsymbol{x}^0 > \mathbf{0}$, 所以 $\lambda_0^m\boldsymbol{x}^0 \gg 0$, 从而 $\boldsymbol{x}^0 \gg \mathbf{0}$. 因为定理 2.1(b) 证明了 $|\lambda| \leqslant \lambda_0$ 只须 $\boldsymbol{A} > \mathbf{0}$. 假设 $|\lambda| = \lambda_0$, 且对某个 $\boldsymbol{z} \neq 0$ 有 $\boldsymbol{A}\boldsymbol{z} = \lambda\boldsymbol{z}$, 则 $\boldsymbol{A}^m\boldsymbol{z} = \lambda^m\boldsymbol{z}$, $\boldsymbol{A}^m\boldsymbol{x}^0 = \lambda_0^m\boldsymbol{x}^0$, 且 $|\lambda^m| = \lambda_0^m$. 如果我们知道, λ_0^m 是 $\boldsymbol{A}^m$ 的最大正特征值, 其证明与定理 2.1 的证明类似. 因为 $\boldsymbol{A}^m \gg \mathbf{0}$, 由定理 2.1 可知 $\boldsymbol{A}^m$ 有一个最大的正特征值, 且对应的特征向量的所有坐标都是正的. 因此, 若 λ_0^m 不是 $\boldsymbol{A}^m$ 的最大正特征值, 则 $\boldsymbol{A}^m$ 有两个正的特征值 $\lambda_1 > \lambda_2$ 且相应的特征向量 $\boldsymbol{x}_1, \boldsymbol{x}_2 \gg \mathbf{0}$. 但这是不可能的; 事实上, 令 $\mu > 0$ 使得 $\boldsymbol{x}_2 - \mu\boldsymbol{x}_1 > \mathbf{0}$ 但不是 $\gg \mathbf{0}$, 则 $\boldsymbol{A}^m(\boldsymbol{x}_2 - \mu\boldsymbol{x}_1) \gg \mathbf{0}$. 另一方面, $\boldsymbol{A}^m(\boldsymbol{x}_2 - \mu\boldsymbol{x}_1) = \lambda_2\boldsymbol{x}_2 - \mu\lambda_1\boldsymbol{x}_1 = \lambda_2(\boldsymbol{x}_2 - \mu\boldsymbol{x}_1) - (\lambda_1 - \lambda_2)\mu\boldsymbol{x}_1$. 因为第一项不是 $\gg \mathbf{0}$, 而第二项是 $\gg 0$, 于是得出矛盾. (c) 的证明与定理 2.1(c) 是一样的, 只要注意到 $\boldsymbol{A}$ 的任意特征向量是 $\boldsymbol{A}^m$ 的一个特征向量. ■

545

为了进一步研究矩阵 $\boldsymbol{A} > \mathbf{0}$, 且对于某些整数 $m > 0$, 使得 $\boldsymbol{A}^m \gg \mathbf{0}$, 我们引入一个秩为 1 的矩阵, 其形式为

$$\boldsymbol{P} = \|x_i^0 f_j^0\|,$$

这里, x^0 同上, $\boldsymbol{f}^0 \gg \mathbf{0}$ 满足 $\boldsymbol{f}^0\boldsymbol{A} = \lambda_0\boldsymbol{f}^0$, 且乘以适当的因子使其正则化, 得 $\sum_{i=1}^{n} x_i^0 f_i^0 = 1$. 则 $\boldsymbol{P}$ 有以下性质:

(i) 对于任意的向量 $\boldsymbol{x}, \boldsymbol{f}$, $\boldsymbol{P}\boldsymbol{x} = (\boldsymbol{x}, \boldsymbol{f}^0)\boldsymbol{x}^0$, $\boldsymbol{f}\boldsymbol{P} = (\boldsymbol{f}, \boldsymbol{x}^0)\boldsymbol{f}^0$, 特别地, $\boldsymbol{P}\boldsymbol{x}^0 = \boldsymbol{x}^0$, $\boldsymbol{f}^0\boldsymbol{P} = \boldsymbol{f}^0$.

(ii) $\boldsymbol{P}^2 = \boldsymbol{P}$.

(iii) $\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{A} = \lambda_0\boldsymbol{P}$.

通过直接计算便可证明前两个结论; 对于第三个结论, 我们知道对于任意的向量 $\boldsymbol{x}$, 都有

$$\boldsymbol{A}\boldsymbol{P}\boldsymbol{x} = \boldsymbol{A}(\boldsymbol{x}, \boldsymbol{f}^0)\boldsymbol{x}^0 = (\boldsymbol{x}, \boldsymbol{f}^0)\boldsymbol{A}\boldsymbol{x}^0 = (\boldsymbol{x}, \boldsymbol{f}^0)\lambda_0\boldsymbol{x}^0 = \lambda_0\boldsymbol{P}\boldsymbol{x},$$

所以 $\boldsymbol{A}\boldsymbol{P} = \lambda_0\boldsymbol{P}$; 类似地, 有 $\boldsymbol{f}\boldsymbol{P}\boldsymbol{A} = \boldsymbol{f}\lambda_0\boldsymbol{P}$, 这就意味着 $\boldsymbol{P}\boldsymbol{A} = \lambda_0\boldsymbol{P}$.

现在, 我们引用以下事实 (不予证明): 设 $\boldsymbol{B}$ 是一个 (方) 阵, 令 $\boldsymbol{B}^n = \|b_{ij}^{(n)}\|$, 且

$$r = \max_{i,j} \overline{\lim_{n\to\infty}} \sqrt[n]{|b_{ij}^{(n)}|}.$$

则 $\boldsymbol{B}$ 有一个特征值 λ^*, 使得 $|\lambda^*| = r$, 若 λ 是 $\boldsymbol{B}$ 的任意其他特征值, 则 $|\lambda| \leqslant r$. 通常, r 被称为 $\boldsymbol{B}$ 的谱半径. 现在, 我们在此基础上证明下面的定理.

定理 2.3 若 $\boldsymbol{A} > 0$, 且存在某一整数 $m > 0$ 使 $\boldsymbol{A}^m >> 0$, λ_0 和 $\boldsymbol{P}$ 如上定义, 则

546

$$\frac{1}{\lambda_0^n}\boldsymbol{A}^n \to \boldsymbol{P}, (n \to \infty).$$

证明 我们可以断言, 若 λ 是矩阵 $\boldsymbol{B} = \boldsymbol{A} - \lambda_0\boldsymbol{P}$ 的特征值, 则 $|\lambda| < \lambda_0$. 事实上, 假定存在向量 $\boldsymbol{z} \neq 0$ 使得 $\boldsymbol{Bz} = \lambda\boldsymbol{z}$, 于是:

$$\lambda\boldsymbol{Pz} = \boldsymbol{PBz} = \boldsymbol{P}(\boldsymbol{A} - \lambda_0\boldsymbol{P})\boldsymbol{z} = (\lambda_0\boldsymbol{P} - \lambda_0\boldsymbol{P}^2)\boldsymbol{z} = \lambda_0(\boldsymbol{P} - \boldsymbol{P})\boldsymbol{z} = 0,$$

因此, 由 $\boldsymbol{Bz} = \lambda\boldsymbol{z}$ 可推得 $\boldsymbol{Az} = \lambda\boldsymbol{z}$. 由定理 2.2 可知, $\lambda = \lambda_0$ 或者 $|\lambda| < \lambda_0$. 若 $\lambda = \lambda_0$, 则 $\boldsymbol{Az} = \lambda_0\boldsymbol{z}$, 从而 $\boldsymbol{z}$ 是 $\boldsymbol{x}_0$ 的倍数. 但是, 如上所述, $\lambda\boldsymbol{Pz} = \lambda\boldsymbol{z} \neq 0$ 是不可能的. 故矩阵 $\boldsymbol{B}$ 的谱半径 r 满足: $r < \lambda_0$.

令 ρ 满足: $r < \rho < \lambda_0$, 因为

$$r = \overline{\lim_{n\to\infty}} \sqrt[n]{\max_{i,j}|b_{ij}^{(n)}|} < \rho,$$

对于充分大的 n, 有 $\max\limits_{i,j}|b_{i,j}^{(n)}| < \rho^n$. 利用 P 的性质 $(ii) - (iii)$, 由归纳法易证

$$\boldsymbol{B}^m = \boldsymbol{A}^m - \lambda_0^m\boldsymbol{P}$$

或者

$$\frac{\boldsymbol{A}^m}{\lambda_0^m} = \frac{\boldsymbol{B}^m}{\lambda_0^m} + \boldsymbol{P}.$$

因为对于充分大的 n, 有 $\max\limits_{i,j}|b_{ij}^{(n)}| < \rho^n$, 所以

$$\left|\frac{b_{ij}^{(m)}}{\lambda_0^m}\right| \leqslant \left(\frac{\rho}{\lambda_0}\right)^m \to 0,$$

从而, $\boldsymbol{B}^m/\lambda^m \to 0$. ■

Frobenius 第二定理 下面是一个主要的 Frobenius 定理.

定理 2.4 设 $A > 0$, 且 λ_0 如定理 2.1 所定义. 则

(a) λ_0 是 A 的特征值, 具有一个特征向量 $x^0 > 0$;

(b) 若 λ 是 A 的任意其他特征值, 则 $|\lambda| \leqslant \lambda_0$;

(c) 当 $x^0 \gg 0$ 时, $\displaystyle\frac{1}{m}\sum_{i=1}^{m}\frac{A^i}{\lambda_0^i}$ 收敛;

(d) 若 λ 是 A 的特征值, 且 $|\lambda| = \lambda_0$, 则 $\eta = \lambda/\lambda_0$ 是一个单位根, 且 $\eta^m\lambda_0$ 是 $\boldsymbol{A}$ 的一个特征值, 其中 $m = 0, 1, 2, 3, \cdots$.

证明 (a) 令 $\boldsymbol{E}$ 是元素全为 1 的矩阵, 故对一切 $\delta>0$, 有 $A+\delta E\gg 0$. 令 $0<\delta_1<\delta_2$, 选择 $\boldsymbol{x}=(x_1,\cdots,x_n)>0$, 使满足 $\sum_{i=1}^{n}x_i=1$. 由 $(\boldsymbol{A}+\delta_1\boldsymbol{E})\boldsymbol{x}\geqslant\lambda\boldsymbol{x}$ 可推得

547

$$(\boldsymbol{A}+\delta_2\boldsymbol{E})\boldsymbol{x}=(\boldsymbol{A}+\delta_1\boldsymbol{E})\boldsymbol{x}+(\delta_2-\delta_1)\boldsymbol{E}\boldsymbol{x}\geqslant[\lambda+(\delta_2-\delta_1)]\boldsymbol{x}.$$

因此, 若 $\lambda_0(\delta)$ 是对应于矩阵 $\boldsymbol{A}+\delta\boldsymbol{E}$ 的 λ_0 值, 可推知 $\lambda_0(\delta)$ 是关于 δ 的递增函数. 注意到, $\lambda_0(0)$ 是对应于 $\boldsymbol{A}$ 本身的 λ_0 值. 根据定理 2.1, 存在一个正规化的向量 $\boldsymbol{x}(\delta)\gg 0$, $\sum_{i=1}^{n}x_i(\delta)=1$, 且满足:

$$(\boldsymbol{A}+\delta\boldsymbol{E})\boldsymbol{x}(\delta)=\lambda_0(\delta)\boldsymbol{x}(\delta).$$

令 $\delta_1>\delta_2>\cdots$ 为一收敛于 0 的正项序列. 根据定理 2.1 的证明, 我们可找到整数 $n_1,n_2,\cdots$, 使得: $\lim\limits_{j\to\infty}\boldsymbol{x}(\delta_{n_j})\to\boldsymbol{x}^0$, 其中向量 $\boldsymbol{x}^0>0$ 满足 $\sum_{i=1}^{n}x_i^0=1$. 显然, $\boldsymbol{A}+\delta_{n_j}\boldsymbol{E}\to\boldsymbol{A}$ 且 $\lambda_0(\delta_{n_j})\to\lambda'\geqslant\lambda_0$. 因为

$$(\boldsymbol{A}+\delta_{n_j}\boldsymbol{E})\boldsymbol{x}(\delta_{n_j})=\lambda_0(\delta_{n_j})\boldsymbol{x}(\delta_{n_j}),$$

令 $j\to\infty$ 可知, $\boldsymbol{A}\boldsymbol{x}^0=\lambda'\boldsymbol{x}^0$. 但根据定理 2.1(b) 所证明的 λ_0 的性质, 知 $\lambda_0\geqslant\lambda'$; 故 $\lambda_0=\lambda'$, 即 (a) 得证.

(b) 的证明等同于 $\boldsymbol{A}\gg 0$ 的情况.

对于 (c) 和 (d) 的证明, 显然, 不失一般性地, 可以假设 $\lambda_0=1$, 否则我们可以把 A 中的每一个元素都除以 λ_0.

(c) 因为 $\boldsymbol{A}\boldsymbol{x}^0=\boldsymbol{x}^0$, 所以 $\boldsymbol{A}^m\boldsymbol{x}^0=\boldsymbol{x}^0$. 把这个等式依分量展开, 易得

$$0\leqslant a_{ij}^{(m)}\leqslant\frac{\max\limits_i x_i^0}{\min\limits_i x_i^0}.$$

故 $\boldsymbol{A}^m$ 的元素是一致有界的.

令 $\boldsymbol{L}=\{\boldsymbol{x}|\boldsymbol{A}\boldsymbol{x}=\boldsymbol{x}\}$ 和 $\boldsymbol{K}=\{\boldsymbol{y}|$ 存在 $\boldsymbol{x}$, 使 $\boldsymbol{y}=(\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}\}$, 即 $\boldsymbol{L}$ 是 $\boldsymbol{A}$ 的固定点的线性空间, $\boldsymbol{K}$ 是矩阵 $\boldsymbol{I}-\boldsymbol{A}$ 的列的线性空间. 另外, 定义

$$\boldsymbol{S}_m=\frac{\boldsymbol{A}+\boldsymbol{A}^2+\cdots+\boldsymbol{A}^m}{m}$$

易见, $\boldsymbol{L}$ 是一个闭线性空间, 从而对每一个 $\boldsymbol{x}\in\boldsymbol{L}$, 成立

$$\boldsymbol{S}_m\boldsymbol{x}=\frac{\boldsymbol{A}+\boldsymbol{A}^2+\cdots+\boldsymbol{A}^m}{m}\boldsymbol{x}=\boldsymbol{x}$$

所以, $\lim_{m\to\infty} \boldsymbol{S}_m\boldsymbol{x} = \boldsymbol{x}$. 我们将要证明, 对任一个 $\boldsymbol{x}\in\boldsymbol{K}$, $\boldsymbol{S}_m\boldsymbol{x}$ 同样收敛, 而且 n 维
548
向量空间中的每一个向量 $\boldsymbol{x}$ 都属于 $\boldsymbol{L}\oplus\boldsymbol{K}$(空间 $\boldsymbol{L}$ 与 $\boldsymbol{K}$ 的直和). 即 (c) 得证.

可断言, 对每一个 $\boldsymbol{y}\in\boldsymbol{K}$, $\lim_{m\to\infty}\boldsymbol{S}_m\boldsymbol{y} = \boldsymbol{0}$. 事实上, 因为存在某个 $\boldsymbol{x}$, 可使 $\boldsymbol{y} = (\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}$, 则推知

$$\boldsymbol{S}_m\boldsymbol{y} \equiv \frac{\boldsymbol{A}\boldsymbol{y}+\boldsymbol{A}^2\boldsymbol{y}+\cdots+\boldsymbol{A}^m\boldsymbol{y}}{m} = \frac{\boldsymbol{A}\boldsymbol{x}-\boldsymbol{A}^{m+1}\boldsymbol{x}}{m}$$

又由于 $\boldsymbol{A}^m$ 的元素是一致有界的, 故当 $m\to\infty$ 时, $\boldsymbol{S}_m\boldsymbol{y}\to 0$.

为了证明任意向量 $\boldsymbol{x}$ 是 $\boldsymbol{L}$ 和 $\boldsymbol{K}$ 中的向量之和. 考虑:

$$\boldsymbol{x} = (\boldsymbol{x}-\boldsymbol{S}_m\boldsymbol{x}) + \boldsymbol{S}_m\boldsymbol{x} = \boldsymbol{y}_m + \boldsymbol{z}_m$$

因为 $\boldsymbol{A}^m$ 的元素是一致有界的, 所以 $\boldsymbol{y}_m$ 和 $\boldsymbol{z}_m$ 的元素也是有界的. 故存在一正整数序列 $m_1 < m_2 < \cdots$, 和一向量 z^0, 使得

$$\lim_{i\to\infty}\boldsymbol{z}_{m_i} = \boldsymbol{z}_0$$

因为

$$\boldsymbol{z}_{m_i} - \boldsymbol{A}\boldsymbol{z}_{m_i} = \frac{\boldsymbol{A}-\boldsymbol{A}^{m_i+1}}{m_i}\boldsymbol{x}\to 0, (i\to\infty),$$

我们有

$$\boldsymbol{z}_0 = \lim_{i\to\infty}\boldsymbol{z}_{m_i} = \lim_{i\to\infty}\boldsymbol{A}\boldsymbol{z}_{m_i} = \boldsymbol{A}\lim_{i\to\infty}\boldsymbol{z}_{m_i} = \boldsymbol{A}\boldsymbol{z}_0,$$

而且 $\boldsymbol{z}_0\in\boldsymbol{L}$.
同样地,

$$\begin{aligned}\boldsymbol{y}_m &= \boldsymbol{x}-\boldsymbol{S}_m\boldsymbol{x} = \frac{1}{m}[(\boldsymbol{x}-\boldsymbol{A}\boldsymbol{x})+(\boldsymbol{x}-\boldsymbol{A}^2\boldsymbol{x})+\cdots+(\boldsymbol{x}-\boldsymbol{A}^m\boldsymbol{x})]\\ &= (\boldsymbol{I}-\boldsymbol{A})\left[\frac{\boldsymbol{x}}{m}+\frac{(\boldsymbol{I}+\boldsymbol{A})\boldsymbol{x}}{m}+\frac{(\boldsymbol{I}+\boldsymbol{A}+\boldsymbol{A}^2)\boldsymbol{x}}{m}+\cdots+\frac{(\boldsymbol{I}+\boldsymbol{A}+\cdots+\boldsymbol{A}^{m-1})\boldsymbol{x}}{m}\right],\end{aligned}$$

故 $\boldsymbol{y}_m\in\boldsymbol{K}$. 因为 $\boldsymbol{K}$ 是闭的线性子空间, $\boldsymbol{y}_m$ 的元素是一致有界的, 所以当 $i\to\infty$ 时, 有 $\boldsymbol{y}_{m_i}\to\boldsymbol{x}-\boldsymbol{z}_0\in\boldsymbol{K}$. 因此, $\boldsymbol{x} = (\boldsymbol{x}-\boldsymbol{z}_0)+\boldsymbol{z}_0$, 其中 $\boldsymbol{x}-\boldsymbol{z}_0\in\boldsymbol{K}, \boldsymbol{z}_0\in\boldsymbol{L}$, 证毕.

(d) 可知存在向量 $\boldsymbol{f}^0 > \boldsymbol{0}$, 满足 $\boldsymbol{f}^0\boldsymbol{A} = \boldsymbol{f}^0$. 首先假设 $\boldsymbol{f}^0 \gg \boldsymbol{0}$. 如果 $\lambda\neq 1$, $|\lambda| = 1$, 且存在 $\boldsymbol{x}\neq\boldsymbol{0}$, 使得 $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$, 则

549

$$\sum_{j=1}^{n} a_{ij}x_j = \lambda x_i, i = 1,2,\cdots,n$$

所以

$$\sum_{j=1}^{n} a_{ij}|x_j| \geqslant |x_i| \quad 或 \quad \boldsymbol{A}|\boldsymbol{x}| \geqslant |\boldsymbol{x}|$$

但若 $\boldsymbol{A}|\boldsymbol{x}| > |\boldsymbol{x}|$, 则 $(\boldsymbol{f}^0, |\boldsymbol{x}|) < (\boldsymbol{f}^0, \boldsymbol{A}|\boldsymbol{x}|) = (\boldsymbol{f}^0\boldsymbol{A}, |\boldsymbol{x}|) = (\boldsymbol{f}^0, |\boldsymbol{x}|)$, 这将错误导致 $\boldsymbol{A}|\boldsymbol{x}| = |\boldsymbol{x}|$. 因此

$$\sum_{j=1}^{n} a_{ij}|x_j| = |x_i| = \left|\sum_{j=1}^{n} a_{ij}x_j\right| \qquad i = 1, 2, \cdots, n$$

故存在常数 $\mu_1, \cdots, \mu_n$, $|\mu_i| = 1$, 满足:

$$(*) \quad a_{ij}x_j = a_{ij}|x_j|\mu_i \quad (\forall i, j)$$

设 $\boldsymbol{x} \cdot \boldsymbol{y}$ 表示向量 $(x_1y_1, \cdots, x_ny_n)'$. 将前面的关系式 (*) 乘以 μ_j^r(即 μ_j 的 r 次幂), 然后对下标 j 求和, 可得:

$$\boldsymbol{A}(\boldsymbol{x} \cdot \boldsymbol{\mu}^r) = \boldsymbol{\mu} \cdot \boldsymbol{A}(|\boldsymbol{x}| \cdot \boldsymbol{\mu}^r).$$

同时, 上式 (*) 对下标 i 求和, 可得:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{\mu} \cdot \boldsymbol{A}|\boldsymbol{x}|,$$

由此可得

$$\lambda\boldsymbol{x} = \boldsymbol{\mu} \cdot |\boldsymbol{x}|.$$

因此

$$\boldsymbol{A}(\boldsymbol{x} \cdot \boldsymbol{\mu}^r) = \boldsymbol{\mu} \cdot \boldsymbol{A}(\lambda\boldsymbol{x} \cdot \boldsymbol{\mu}^{r-1}) = \lambda\boldsymbol{\mu} \cdot \boldsymbol{A}(\boldsymbol{x} \cdot \boldsymbol{\mu}^{r-1}), \quad r = 1, 2, \cdots,$$

由此递归式可推得: $\boldsymbol{A}(\boldsymbol{x} \cdot \boldsymbol{\mu}^r) = \lambda^{r+1}(\boldsymbol{\mu}^r \cdot \boldsymbol{x})$. 故 λ^r 是 $\boldsymbol{A}$ 的特征值, $r = 1, 2, \cdots$. 因为 $\boldsymbol{A}$ 的特征值的个数是有限的, 所以 λ 必为单位根.

现假设 $\boldsymbol{f}^0 > \boldsymbol{0}$ 而不满足 $\boldsymbol{f}^0 \gg \boldsymbol{0}$. 如果必要的话, 通过重新排列 $\boldsymbol{A}$ 的行与列, 可以假设 $\boldsymbol{f}^0 = (f_1^0, \cdots, f_r^0, 0, \cdots, 0)$, 其中 $f_i^0 > 0, i = 1, \cdots, r$. 因为 $\boldsymbol{A} > \boldsymbol{0}$, 由关系式 $\boldsymbol{f}^0\boldsymbol{A} = \boldsymbol{f}^0$ 可写出分块阵

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_1 & \boldsymbol{0} \\ \boldsymbol{B} & \boldsymbol{A}_2 \end{pmatrix},$$

其中 $\boldsymbol{A}_1$ 是一个 $r \times r$ 阶矩阵, 而 $\boldsymbol{A}_2$ 是一个 $(n-r) \times (n-r)$ 阶矩阵. 而且向量 $(f_1^0, \cdots, f_r^0)$ 是 $\boldsymbol{A}_1$ 的左特征向量, 相应的特征值为 1. 令 λ 是 $\boldsymbol{A}$ 的特征值. 如果 λ 同时也是 $\boldsymbol{A}_1$ 的一个特征值, 那么用 $\boldsymbol{A}_1$ 替换 $\boldsymbol{A}$, 我们又回到了原来考虑的问题中. 如果 λ 不是 $\boldsymbol{A}_1$ 的特征值, 那么它必是 $\boldsymbol{A}_2$ 的特征值. 但 $\boldsymbol{A}_2$ 的特征值一定是 $\boldsymbol{A}$ 的特征值, 且绝对值不超过 1. 同时, 因为 $\boldsymbol{A}_2 > 0$, 故它有一个最大的正特征值, 即所有特征值的绝对值的上界. 由于 $|\lambda| = 1$, $\boldsymbol{A}_2$ 最大的正特征值正是 1. 显然, 我们可 550

以对 $\boldsymbol{A}_2$ 应用前面的分析. 因此它或者有一个左特征向量 $\gg 0$, 其相应的特征值为 1, 或者它会有如下形式 (可适当重排行与列):

$$\boldsymbol{A}_2 = \begin{pmatrix} \boldsymbol{A}_3 & \boldsymbol{0} \\ \boldsymbol{B}_1 & \boldsymbol{A}_4 \end{pmatrix}.$$

只要有限步地重复这一过程, 我们可以将问题简化为存在一个左特征向量 $\gg \boldsymbol{0}$, 对应特征值为 1 的情况. ■

下面的推论将产生一些关于正矩阵 A 的谱半径 $\lambda_0(A)$ 的有用信息. 第一个推论只是定理 2.4 中结论 (a) 和 (b) 的简单重述.

推论 2.1 若 $\boldsymbol{A} > \boldsymbol{0}$, 则 $\boldsymbol{A}$ 的最大特征值 $\lambda_0 = \lambda_0(\boldsymbol{A})$ 是实非负的, 而且可以表示成 $\lambda_0 = \max_\Lambda \lambda$, 其中

$$\Lambda = \{\lambda | \boldsymbol{A}\boldsymbol{x} \geqslant \lambda \boldsymbol{x},\ \text{存在}\ \boldsymbol{x} > 0\}.$$

推论 2.2 若 $\boldsymbol{A} > \boldsymbol{0}$, 且存在 $\boldsymbol{x}^0 \gg \boldsymbol{0}$ 使得 $\boldsymbol{A}\boldsymbol{x}^0 \leqslant \mu \boldsymbol{x}^0$, 则 μ 是 $\lambda_0(\boldsymbol{A})$ 的一个上界.

证明 将 $\boldsymbol{A}$ 乘到 $\boldsymbol{A}\boldsymbol{x}^0 \leqslant \mu \boldsymbol{x}^0$ 的两边, 我们得到 $\boldsymbol{A}^2\boldsymbol{x}^0 \leqslant \mu \boldsymbol{A}\boldsymbol{x}^0 \leqslant \mu^2 \boldsymbol{x}^0$, 重复这一步骤可得

$$\boldsymbol{A}^n \boldsymbol{x}^0 \leqslant \mu^n \boldsymbol{x}^0, \quad n = 1, 2, \cdots$$

由此易得

$$a_{ij}^{(n)} \leqslant \mu^n \frac{\max\limits_i x_i^0}{\min\limits_i x_i^0},$$

所以

$$\lambda_0(\boldsymbol{A}) = \overline{\lim_{n\to\infty}} \sqrt[n]{\max_{i,j} |a_{ij}^{(n)}|} \leqslant \mu.$$ ■

推论 2.3 若 $\boldsymbol{A} \geqslant \boldsymbol{B} \geqslant \boldsymbol{0}$, 则 $\lambda_0(\boldsymbol{B}) \leqslant \lambda_0(\boldsymbol{A})$.

证明 可由推论 2.1 推得, 或者由关系式

$$\lambda_0(\boldsymbol{A}) = \overline{\lim_{n\to\infty}} \sqrt[n]{\max_{i,j} a_{ij}^{(n)}} \geqslant \overline{\lim_{n\to\infty}} \sqrt[n]{\max_{i,j} b_{ij}^{(n)}} = \lambda_0(\boldsymbol{B}),$$

551 显然, 由 $\boldsymbol{A} \geqslant \boldsymbol{B} \geqslant \boldsymbol{0}$, 可推得 $\boldsymbol{A}^n \geqslant \boldsymbol{B}^n \geqslant \boldsymbol{0}$, $n = 1, 2, \cdots$. ■

索 引

索引中页码为英文原书页码，与书中页边标注的页码一致。

C

D

E

Y

前言